



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 1022784

To: Sheridan Swope
Location: CM1-10D01
Art Unit: 1652
Wednesday, August 27, 2003

Case Serial Number: 09/696872

From: Beverly Shears
Location: Biotech-Chem Library
CM1-1E05
Phone: 308-4994

beverly.shears@uspto.gov

Search Notes

This Page Blank (uspto)

From: Swope, Sheridan
Sent: Monday, August 25, 2003 6:21 PM
To: STIC-Biotech/ChemLib
Subject: FW: 09696872

-----Original Message-----

From: Chan, Christina
Sent: Monday, August 25, 2003 6:20 PM
To: Swope, Sheridan; STIC-Biotech/ChemLib
Subject: RE: 09696872

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

-----Original Message-----

From: Swope, Sheridan
Sent: Monday, August 25, 2003 5:18 PM
To: Chan, Christina
Subject: 09696872

May I have this rushed?

For 09696872, pls search:

SID 23 against the AA and NT data bases.

SID 24 against the AA and NT data bases.

SID 37 against the AA and NT data bases.

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
sheridan.swope@uspto.gov
703-305-1696 (voice)
703-308-3014 (FAX)
Mailbox: CM1 Rm10D01
Office: CM1 Rm12D12

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 08-27-03
Searcher: Beverly 24994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: _____

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
☒ Other CGN

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 16:08:15 / Search time 4834.58 Seconds
(without alignments)

922.344 Million cell updates/sec

Title: US-09-696-872-23

Sequence: 1 MRVMTGLALAAVCSAAK.....PKPQKPEPEGTGSSEKDEL 109

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USFO.spool/US09696872/runat_26082003_151137_3223/app_query.fasta_1.462
-DB=genEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=ppct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09696872 @CGN 1 1 3608 @runat_26082003_151137_3223 -NCPU=6 -ICPU=3
-NO MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rdi:*
36: em_htg_mam:*
37: em_htg_vir:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	579	99.1	372	6	AR121628 Sequence
2	459	78.6	372	6	AR121631 Sequence
3	429	73.5	372	6	AR121630 Sequence
4	428.5	73.4	387	6	AR121624 Sequence
5	422	72.3	444	6	AR121632 Sequence
6	414.5	71.0	387	6	AR121623 Sequence
7	371.5	63.6	369	6	AR121627 Sequence
8	345.5	59.2	369	6	AR121626 Sequence
9	337.5	57.8	357	6	AR121625 Sequence
10	304.5	52.1	315	6	AR121629 Sequence
11	262.5	44.9	2274	6	AB086984 Homo sapi
12	262.5	44.9	2439	6	AX329932 Sequence
13	262.5	44.9	2439	6	AX330698 Sequence
14	262.5	44.9	2439	6	AX330917 Sequence
15	262.5	44.9	2439	6	HUMCOMP
16	266.5	43.9	764	6	BD126595
17	266.5	43.9	764	6	BD126165
18	256.5	43.9	1779	6	BD127132
19	256.5	43.9	1779	6	AK074508 Homo sapi
20	247.5	42.4	2302	4	AF325902
21	245	42.0	305	12	EVER291687
22	236	40.4	2438	10	AF033530
23	234	40.1	2421	10	RNCOMP
24	227.5	39.0	2710	9	BC033676
25	210	36.0	323	12	EVER291685
26	155	26.5	302	12	EVER291686
27	154.5	26.5	46275	9	AC0003107
28	153	26.2	4469	9	AB018333
29	153	26.2	4744	9	AK025495
30	153	26.2	154288	9	AL513164
31	153	26.2	170573	2	AC138890
32	149	25.5	144	6	AX39800
33	149	25.5	177	6	A39798
34	146.5	25.1	3160	10	AF102887
35	146.5	25.1	3365	9	BC050456
36	146	25.0	8724	4	CDR311945
37	145.5	24.9	3074	6	AX094825
38	145.5	24.9	3074	6	AX336954
39	145.5	24.9	3074	6	AX480839
40	145.5	24.9	3074	9	HSTR0M4
41	144.5	24.7	3123	10	RNTSP4
42	144	24.3	78	6	A92276
43	142	24.3	85877	2	AC091343
44	142	24.3	244328	7	AC094938
45	141	24.1	210148	10	AC122864

RESULT 1

ALIGNMENTS

AR121628 Sequence
AR121631 Sequence
AR121630 Sequence
AR121624 Sequence
AR121632 Sequence
AR121623 Sequence
AR121627 Sequence
AR121626 Sequence
AR121625 Sequence
AR121629 Sequence
AB086984 Homo sapi
AX329932 Sequence
AX330698 Sequence
AX330917 Sequence
I32137 Human germ
BD126595 Primer fo
BD126165 Primer fo
BD127132 Primer fo
AK074508 Homo sapi
AF325902 Equus cab
AJ291687 Expressio
AF033530 Mus muscu
X72914 R.norvegicu
BC033676 Homo sapi
AJ291685 Expressio
AJ291686 Expressio
AC0003107 Human DNA
AB018333 Homo sapi
AK025495 Homo sapi
AL513164 Human DNA
AC138890 Homo sapi
AX39800 Sequence 34
A39798 Sequence 32
AF102887 Mus muscu
BC050456 Homo sapi
AJ131945 Camelus d
AX094825 Sequence
AX336954 Sequence
AX480839 Sequence
Z19585 H.sapiens m
X89963 R.norvegicu
A92276 Sequence 13
Continuation (8 of
AC094938 Rattus no
AC122864 Mus muscu

AR121628 LOCUS AR121628 372 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 24 from patent US 6160088.
 ACCESSION AR121628
 VERSION AR121628.1 GI:14105204
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 372)
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 24 12-DEC-2000;
 FEATURES Location/Qualifiers
 source 1..372
 BASE COUNT 91 a 102 c 122 g 57 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4,47e-45 Length: 372
 Score: 579.00 Matches: 108
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.14% Indels: 0
 DB: Gaps: 0

US-09-696-872-23 (1-109) x AR121628 (1-372)

OY 2 ArgTYrMetIIleuGlyLeuLeuAlaLeuAlaValaCySerAlaAlaIleuArgGly 21
 DB 16 AGGTACATGATTTTATGGCTTGCTCCCTTCGCGACGCTCGACGCGTCCCAAAAAGGA 75
 OY 22 SerSerLeuGlyGlyAAspCySeSerAspLeuGlyProGlnMetLeuArgGluLeuGln 41
 DB 76 TCCAGCTGGGTGGAGACTGTGTTCAGACCTGGGCGCGACAGATCTTCGGGAACCTGCG 135
 OY 42 GluThrAenAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnValaArgGluIle 61
 DB 136 GAATTAACCAACTCTCGGAGAGGTGAAGACCTTCGAGACAGCAGGTTAAAGGAACA 195
 OY 62 ThrPheLeuIysAenThrValMetGluCyAspAlaCySerGlyProGlnProGlnProlys 81
 DB 196 ACGTTCCTGAATAAACACCGGTGATGAGTGTGACGCGTGGGCGCGCCAGCCGACCGGAAA 255
 OY 82 ProGlnProGlnProGlnProGlnProlysProGlnProGlnProGluProGluGlyThr 101
 DB 256 CCGACCGCGCAGCCCGCAGCCGACCGGAAACCGCAGCCGAAACCGGAAGGTACC 315
 OY 102 GlySerSerGluIysAspGluLeu 109
 DB 316 GGATCATCAGAAAAAGATGAGTTG 339

RESULT 2
 AR121631 LOCUS AR121631 372 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 30 from patent US 6160088.
 ACCESSION AR121631
 VERSION AR121631.1 GI:14105207
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 372)
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 30 12-DEC-2000;
 FEATURES Location/Qualifiers
 source 1..372
 BASE COUNT 102 a 96 c 106 g 68 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5.29e-34 Length: 372
 Score: 459.00 Matches: 84
 Percent Similarity: 86.11% Conservative: 9
 Best Local Similarity: 77.78% Mismatches: 15
 Query Match: 78.60% Indels: 0
 DB: Gaps: 0

US-09-696-872-23 (1-109) x AR121631 (1-372)

OY 2 ArgTYrMetIIleuGlyLeuLeuAlaLeuAlaValaCySerAlaAlaIleuArgGly 21
 DB 16 AGGTACATGATTTTATGGCTTGCTCCCTTCGCGACGCTCGACGCGTCCCAAAAAGGA 75
 OY 22 SerSerLeuGlyGlyAAspCySeSerAspLeuGlyProGlnMetLeuArgGluLeuGln 41
 DB 76 TCCAGCTGGGTGGAGACTGTGTTCAGACCTGGGCGCGACAGATCTTCGGGAACCTGCG 135
 OY 42 GluThrAenAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnValaArgGluIle 61
 DB 136 CAATTAACCAACTCTCGGAGAGGTGAAGACCTTCGAGACAGCAGGTTAAAGGAACA 195
 OY 62 ThrPheLeuIysAenThrValMetGluCyAspAlaCySerGlyProGlnProGlnProlys 81
 DB 196 TCATTTTTCGAAACACCATGATGCTGAATGCCAGGCTTGGGTCCGACGCGCAGCCGAAA 255
 OY 82 ProGlnProGlnProGlnProGlnProlysProGlnProGlnProGluProGluGlyThr 101
 DB 256 CCGACCGCGCAGCCCGCAGCCCGACCGGAAACCGCAGCCGAAACCGGAAGGTACC 315
 OY 102 GlySerSerGluIysAspGluLeu 109
 DB 316 GGATCATCAGAAAAAGATGAGTTG 339

RESULT 3
 AR121630 LOCUS AR121630 372 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 28 from patent US 6160088.
 ACCESSION AR121630
 VERSION AR121630.1 GI:14105206
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 372)
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 28 12-DEC-2000;
 FEATURES Location/Qualifiers
 source 1..372
 BASE COUNT 96 a 105 c 109 g 62 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.1e-31 Length: 372
 Score: 429.00 Matches: 78
 Percent Similarity: 84.26% Conservative: 13
 Best Local Similarity: 72.22% Mismatches: 17
 Query Match: 73.46% Indels: 0
 DB: Gaps: 0

US-09-696-872-23 (1-109) x AR121630 (1-372)

OY 2 ArgTYrMetIIleuGlyLeuLeuAlaLeuAlaValaCySerAlaAlaIleuArgGly 21
 DB 16 AGGTACATGATTTTATGGCTTGCTCCCTTCGCGACGCTCGACGCGTCCCAAAAAGGA 75
 OY 22 SerSerLeuGlyGlyAAspCySeSerAspLeuGlyProGlnMetLeuArgGluLeuGln 41
 DB 76 TCCAGCTGGGTGGAGACTGTGTTCAGACCTGGGCGCGACAGATCTTCGGGAACCTGCG 135
 OY 42 GluThrAenAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnValaArgGluIle 61

[illegible]

ACCESSION	AR121632
VERSION	AR121632.1 GI:14105208
KEYWORDS:	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 444)
TITLE	Rothman,J.E., Mayhew,M. and Hoe,M.H.
JOURNAL	KOBL receptor inhibitors
FEATURES	Patent: US 6160088-A 35 12-DEC-2000;
source	Location/Qualifiers
	1..444
BASE COUNT	/organism="unknown"
ORIGIN	115 a 122 c 134 g 73 t
Alignment Scores:	
Pred. No.:	1,65e-30 Length: 444
Score:	422.00 Matches: 82
Percent Similarity:	84.85% Conservative: 2
Best Local Similarity:	82.83% Mismatches: 1
Query Match:	72.26% Indels: 14
DB:	Gaps: 2
US-09-696-872-23 (1-109) x AR121632 (1-444)	
Dn	21 GlySerSerLeuGlyGlyAspCysCyseSerAspLeuGlyProGlnMetLeuArgGluIleu 40
Dn	127 GGATCCAGCGCTGGTGAGAAGTGTGT-----CCACAGATGCTTCGGAATC 174
Dn	41 GlnGluThrAnaAlaLeuGlnAspValArgAspTrpLeuArgGlnValArgGlu 60
Dn	175 CAGAGAGCATATCGCGCGCTGCACAAGCGTAGAAGCTCTTCGACAGCAAGGTCAAGAG 234
Dn	61 IleThrPheLeuLysAsnThrValMetGluCysAspAlaCysGly----- 75
Dn	235 ATCACCTTCCTGAAAGATTACGATGATGATGATACCGTTGCCGAATGCACCCCGCAGCC 294
Dn	76 -----ProGlnProGlnProLysProGlnProGlnProGlnProGlnPro 90
Dn	295 ACCCCCGGACTACTCCGCGACGCCGACCGCAAAACCGACGCCGACGCCGACGCCG 354
Dn	91 LysProGlnProLysProGlnProGlnGlyThrlGlySerSerGluLysAspGluLeu 109
Dn	355 AAACCGCAGCCGAAACCGGAACCGGAAGGTATCCGATCATCAAAAAAGATGAGTTG 411
RESULT 6	
LOCUS	AR121623 387 bp DNA linear PAT 16-MAY-2001
DEFINITION	Sequence 14 from patent US 6160088.
ACCESSION	AR121623
VERSION	AR121623.1 GI:14105199
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 387)
TITLE	Rothman,J.E., Mayhew,M. and Hoe,M.H.
JOURNAL	KOBL receptor inhibitors
FEATURES	Patent: US 6160088-A 14 12-DEC-2000;
source	Location/Qualifiers
	1..387
BASE COUNT	/organism="unknown"
ORIGIN	94 a 110 c 125 g 58 t
Alignment Scores:	
Pred. No.:	7.05e-30 Length: 387
Score:	414.50 Matches: 86
Percent Similarity:	79.46% Conservative: 3
Best Local Similarity:	76.79% Mismatches: 6
Query Match:	70.98% Indels: 17
DB:	Gaps: 3

US-09-696-872-23 (1-109) x AR121623 (1-387)

QY 8 LeuLeuAlaLeuAlaAlaValCySerSerAlaAlaAlaLysGlySerSerLeuGlyGlyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGTG-----CGGCCGAGAGATCCAGAGATCATATCGCGCTG 87

QY 28 CysCySerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
DB 88 -----GACCTAGCCCCACAGATGCTTGAGAACTCCAGAGATCATATCGCGCTG 138

QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 139 CAAAGCGAGAGAGAGCTCTTGAGACAGACAGTCAGAGATCATCTCTTGAAAGATACG 198

QY 68 ValMetGluCyAspAlaCyAspGly-----ProGln 77
DB 199 GTGATGAAATGTGACGCTTGCGGAATGCAGCCCGACGACCCCGGTAATGATCGCGAG 258

QY 78 ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLysProGln 97
DB 259 CCGGACCGGAAACCGCAGCGCGCAGCGCGGAAACCGCAGCGGAAACCGGAAACCGGAA 318

QY 98 ProGlnGlyThrGlySerSerGlyLysAspGluLeu 109
DB 319 CCGGAAAGTACCGGATCATCAGAAAAAGATGAGTTG 354

RESULT 7
AR121627 369 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 22 from patent US 6160088.
ACCESSION AR121627
VERSION AR121627.1 GI:14105203
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 369)
AUTHORS Rochman,J.E., Mayhew,M. and Hoe,M.H.
TITLE KDEL receptor inhibitors
JOURNAL Patent: US 6160088-A 22-12-DEC-2000;
FEATURES
source location/Qualifiers
BASE COUNT 94 a 94 c 121 g 60 t
ORIGIN

Alignment Scores:
Pred. No.: 6.23e-26 Length: 369
Score: 371.50 Matches: 68
Percent Similarity: 79.41% Conservative: 13
Best Local Similarity: 66.67% Mismatches: 18
Query Match: 63.61% Indels: 3
DB: 6 Gaps: 1

US-09-696-872-23 (1-109) x AR121627 (1-369)

QY 8 LeuLeuAlaLeuAlaAlaValCySerSerAlaAlaAlaLysGlySerSerLeuGlyGlyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGTG-----CGGCCGAGAGATCCAGAGATCATATCGCGCTG 90

QY 28 CysCySerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
DB 91 TGTGTGGTGGAGCGACAGACAGATTGATGCGCAGATTAACCAATGATCATATCGCTG 150

QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 151 GGAGGCTCCGAGATGTCAGTGAACAGCAGCGTGAAGAACCATGATCTTCTGAAGAACCC 210

QY 68 ValMetGluCyAspAlaCyAspGlyProGlnProGlnProLysProGlnProGlnProGln 87
DB 211 ATTGCAAGATGCGAGCCCTGTGGCCCGCAGCCGACCGAAACCGCAGCCGACCGCAG 270

QY 88 ProGlnProLysProGlnProLysProGlnProGlnGlyThrGlySerSerGlyLysAsp 107
DB 271 CCGGACCGGAAACCGCAGCGCGCAGAACCGGAACCGGAAGGTACCGGATCATCAGAAAAAGAT 330

QY 108 GluLeu 109
DB 331 GAGTTG 336

RESULT 8
AR121626 369 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 20 from patent US 6160088.
ACCESSION AR121626
VERSION AR121626.1 GI:14105202
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 369)
AUTHORS Rochman,J.E., Mayhew,M. and Hoe,M.H.
TITLE KDEL receptor inhibitors
JOURNAL Patent: US 6160088-A 20-12-DEC-2000;
FEATURES
source location/Qualifiers
BASE COUNT 87 a 103 c 120 g 59 t
ORIGIN

Alignment Scores:
Pred. No.: 1.56e-23 Length: 369
Score: 345.50 Matches: 64
Percent Similarity: 76.47% Conservative: 14
Best Local Similarity: 62.75% Mismatches: 21
Query Match: 59.16% Indels: 3
DB: 6 Gaps: 1

US-09-696-872-23 (1-109) x AR121626 (1-369)

QY 8 LeuLeuAlaLeuAlaAlaValCySerSerAlaAlaAlaLysGlySerSerLeuGlyGlyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGTG-----CGGCCGAGAGATCCAGCTGCTGAGAGAC 90

QY 28 CysCySerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
DB 91 TGTGTGGGAGCGACAGACAGCATTTGTCACCCAGCTCACCTCTTCACACAGATCCTA 150

QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 151 GTGAGCTTCGGAGACGATCCGAGACAGGTGAAGAAATGTCATCATCGGAAACCC 210

QY 68 ValMetGluCyAspAlaCyAspGlyProGlnProGlnProLysProGlnProGlnProGln 87
DB 211 ATCATGAGATTCAGAGTGTGCTCCGACCGCAGCCGAAACCGCAGCCGACCGCAG 270

QY 88 ProGlnProLysProGlnProLysProGlnProGlnGlyThrGlySerSerGlyLysAsp 107
DB 271 CCGGACCGGAAACCGCAGCGCGCAGAACCGGAACCGGAAGGTACCGGATCATCAGAAAAAGAT 330

QY 108 GluLeu 109
DB 331 GAGTTG 336

RESULT 9
AR121625 357 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 18 from patent US 6160088.
ACCESSION AR121625
VERSION AR121625.1 GI:14105201
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 357)

AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 18 12-DEC-2000;
 FEATURES location/Qualifiers
 source 1..357
 /organism="unknown"
 BASE COUNT 84 a 100 c 114 g 59 t
 ORIGIN

Alignment Scores:

Pred. No.: 8.25e-23 Length: 357
 Score: 337.50 Matches: 65
 Percent Similarity: 75.49% Conservative: 12
 Best Local Similarity: 63.73% Mismatches: 18
 Query Match: 57.79% Indels: 7
 DB: 6 Gaps: 2

US-09-696-872-23 (1-109) x AR121625 (1-357)

Qy 8 LeuLeuAlaLeuAlaAlaValaCyseSerAlaAlaAlaLysLysGlySerSerLeuGlyGlyYasp 27
 Db 40 TTGCTGCTGCTGGGCGCGCTG-----CGGGCCGAGGATCCAGCTGGGTGAGAC 90
 Qy 28 CysCysSerAspLeuGlyProGlnMetLeuArgLueGlnGluThrAsnAlaAlaLeu 47
 Db 91 TGTGTAAAGGCGATTG-----GTCAACCCAGCTCACCTCTTCAACACAGATCCTA 138
 Qy 48 GlnAspValaGAspTrpLeuArgGlnGlnValaArgGluIleThrPheLeuLysAsnThr 67
 Db 139 GTGAGCGCTTCGGAGCAGCATCCGAGACCGAGGTAAGGAAATGTCATCATCCGGAACCC 198
 Qy 68 ValMetGluCyAspAlaCyGlyProGlnProGlnProGlnProGlnProGlnProGln 87
 Db 199 ATCATGAGTTCAGAGTGTGGGTCGCGACCGACCGGAAACCGCAGCCGACCGCGAG 258
 Qy 88 ProGlnProLysProGlnProLysProGlnProGlnGlyThrGlySerSerGluLysAsp 107
 Db 259 CCGAGCGCGAAACCGCAGCCGAAACCGGAAACCGGAAAGTACCGGATCATCAGAAAAAGAT 318
 Qy 108 GluLeu 109
 Db 319 GAGTTG 324

RESULT 10 AR121629 315 bp DNA linear PAT 16-MAY-2001
 LOCUS AR121629 Sequence 26 from patent US 6160088.
 DEFINITION AR121629
 ACCESSION AR121629
 VERSION AR121629.1 GI:14105205
 KEYWORDS

SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 315)
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 26 12-DEC-2000;
 FEATURES Location/Qualifiers
 source 1..315
 /organism="unknown"

BASE COUNT 81 a 85 c 79 g 70 t
 ORIGIN

Alignment Scores:
 Pred. No.: 8.03e-20 Length: 315
 Score: 304.50 Matches: 70
 Percent Similarity: 66.67% Conservative: 14
 Best Local Similarity: 63.06% Mismatches: 13
 Query Match: 52.14% Indels: 25
 DB: 6 Gaps: 2

US-09-696-872-23 (1-109) x AR121629 (1-315)

Qy 2 ArgTyMetIleLeuGlyLeuLeuAlaAlaValaCyseSerAlaAlaLysLysGly 21
 Db 16 AGTACATGATTTTAAAGCTTGCTGCGCTTCGGCATCTGCAGCGCTGCCAAAAAGGA 75
 Qy 22 SerSerLeuGlyGlyYaspCyseSerAspLeuGlyProGlnMetLeuArgGluLeuGln 41
 Db 76 TCCAGCTGGGTGAGAGCTGTTCGAAAGCTACA-GAATCTATTTCATATTTCTCTCT 134
 Qy 42 GluThrAsnAlaAlaLeuGlnAspVal-----ArgAsp-TPLeuArgGlnGlnVala 59
 Db 135 CATTTAATATGTCTTCTGCTGATCTATCATCTGATGATCTCTC----- 180
 Qy 59 gGluIleThrPheLeuLysAsnThrValMetGluCyAspAlaCyseGlyProGlnProG 79
 Db 181 -----CCGAGCGCGCA 191
 Qy 79 nProLysProGlnProGlnProGlnProGlnProLysProGlnProGlnProGln 99
 Db 192 GCCGAAACCCGAGCCGAGCCGAGCCGAGCCGAAACCCGAGCCGAAACCGGAA 251
 Qy 99 uGlyThrGlySerSerGlyLysAspGluLeu 109
 Db 252 AGTACCGGATCATCAGAAAAAGATGAGTTG 282
 RESULT 11
 LOCUS AB086984 2274 bp mRNA linear PRI 07-JAN-2003
 DEFINITION Homo sapiens comp mRNA for cartilage oligomeric matrix protein.
 ACCESSION AB086984
 VERSION AB086984
 KEYWORDS AB086984.1 GI:27531065
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Hashimoto,Y. and Mori,H.
 TITLE Human comp cDNA with 5 SNIPS
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2274)
 AUTHORS Hashimoto,Y. and Mori,H.
 TITLE Direct Submission
 JOURNAL Submitted (27-JUN-2002) Hiroshi Mori, Osaka City University Medical
 School, Neuroscience, 1-4-3 Asahimachi, Abeno-ku, Osaka 545-8585,
 Japan (E-mail: mori@med.osaka-cu.ac.jp, Tel:81-6-6645-3920,
 Fax:81-6-6645-3922)
 FEATURES Location/Qualifiers
 source 1..2274
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_type="fibroblast"
 1..2274
 /gene="comp"
 1..2274
 /gene="comp"
 /codon_start=1
 /product="cartilage oligomeric matrix protein"
 /protein_id="BAC53888.1"
 /db_xref="GI:27531066"
 /translation="MVPDTRACVLLTLTALGASGGGSPISGLDGMPLBELQETNA
 LQVRDMLRQVREITFLPKTWTWECDAKQGSVYRGLPSVRLHLCARFCPCPGYAC
 LQTESGRCGCPAGFTGNSHCTDVNECAHPCFPRVCTINTSPGRRCACPCPGYSG
 PTQGVGLAPAKANKQVCTDINECTGQHNCAVNSVCINTRGSFQCGPCPGFYGDGA
 SGCORGAQRCPCPDGSPBSBCEHADCVLERDGRSCVCYRVGMAANGILCGSDTLDFP
 DEKLRCEPCQCRDNCTVYNSGOEVDKIGDACDPDADGDPVNEKXNCPYVNP
 DQNTBDEKWKGDACNDRSQNDOKDQTDGGRGACDDPIDGRIIRNADNCPVNP
 SDKSDSGDGTGDCNCPKSNPDQADVDHDPVGCACSDQDQDSDGSDSNCT
 VPSAQSDESDQGDACDDDDNDNDGVPISRDNCRLVPNPGQDADBDGIEDVCGDQF
 DADKVVYKLDVCEPNAEVLITDFRAFQTVLDEGAGQIDPNVNVINQGEIYQTNMS
 DPGIAGVYTAENGVDPEGFHVNVTVDYDAGFI FGYSOSSSFYVVMKQMEQITWQA
 NPFRVAVEPGIQLKAVKSTGPGEQLRNALMHTGDTESQVRLIMKDPKRNWMDKXSY

gene
 CDS

```

RMFLQHRPOVGIRVRFEYGEPLVADSNVVLDTTMRGRLGVFCPSQENIIMANLRYR
CNDTRPDEYETHQLROQA"
variation 150
/ gene="comp"
/ note="SNIPs of # 1, 2, 3 are localized at the pentamer
formation domain"
/ replace="g"
151
variation / gene="comp"
/ note="SNIPs of # 1, 2, 3 are localized at the pentamer
formation domain"
/ replace="c"
152
variation / gene="comp"
/ note="SNIPs of # 1, 2, 3 are localized at the pentamer
formation domain"
/ replace="c"
326
variation / gene="comp"
/ note="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/ replace="c"
327
variation / gene="comp"
/ note="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/ replace="g"
669
variation / gene="comp"
/ note="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/ replace="g"
670
variation / gene="comp"
/ note="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/ replace="c"
766
variation / gene="comp"
/ note="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/ replace="c"
767
variation / gene="comp"
/ note="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/ replace="g"
854
variation / gene="comp"
/ note="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/ replace="c"
855
variation / gene="comp"
/ note="SNIPs of # 10, 11 are localized in a stuffed
sequence between the EGF-like repeats and the TSP typed
repeats"
/ replace="c"
BASE COUNT 467 a 707 c 749 g 351 t
ORIGIN

Alignment Scores:
Pred. No.: 4,78e-15 Length: 2274
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 9 Gaps: 1

US-09-696-872-23 (1-109) x AB086984 (1-2274)
QY 8 LeuLeuAlaLeuAlaAlaValCySerAlaAlaAlaLysGlySerSerLeuGlyAsp 27

```

```

||||| ||||| : : : : : ||| |||||
Db CTGCTCACCCCTGCTGCGCGCGCGCGAGACAGCGCCGCTGGGC----- 81
QY 28 CysCySerSerAsePleuAglProGlnMerLeuAargGluLeuGlnGluThrAAsnAlaLeu 47
Db 82 -----TCAGACCTGGGGCGCCAGATGCTTGGGAACCTGCGGAAACCAAGCGCGCTG 135
QY 48 GlnAseValAArgAsePTrPLeuAargGlnGlnValAargGluLeuThrPheLeuLysAAsnThr 67
Db 136 CAGGACGTGGCGGAGCTGGCTCGGACAGAGTCAAGGAGATCAAGTCTCTAATAAACCG 195
QY 68 ValMetGluCyAseAAlaCyGlyProGlnProGlnProGlnProGlnProGln 87
Db 196 GTGATGAGTGTGACGCGCTGCGGATGACAGTCAAGTACGAC-CCGCTTACCAGCGT 254
QY 88 ProGlnProLysPro 92
Db 255 GCGGCCCTGCTCCA 269

RESULT 12
AX329932 2439 bp DNA linear PAT 09-JAN-2002
LOCUS
DEFINITION Sequence 441 from Patent WO0194629.
ACCESSION AX329932
VERSION AX329932.1 GI:18102910
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horriگان, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 441 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..2439
location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 503 a 758 c 809 g 369 t
ORIGIN

Alignment Scores:
Pred. No.: 5,14e-15 Length: 2439
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 6 Gaps: 1

US-09-696-872-23 (1-109) x AX329932 (1-2439)
QY 8 LeuLeuAlaLeuAlaAlaValCySerAlaAlaAlaLysGlySerSerLeuGlyAsp 27
Db 53 CTGCTCACCCCTGCTGCGCGCGCGCGAGACAGCGCCGCTGGGC----- 106
QY 28 CysCySerSerAsePleuAglProGlnMerLeuAargGluLeuGlnGluThrAAsnAlaLeu 47
Db 107 -----TCAGACCTGGGGCGCCAGATGCTTGGGAACCTGCGGAAACCAAGCGCGCGCTG 160
QY 48 GlnAseValAArgAsePTrPLeuAargGlnGlnValAargGluLeuThrPheLeuLysAAsnThr 67
Db 161 CAGGACGTGGCGGAGCTGGCTCGGACAGAGTCAAGGAGATCAAGTCTCTAATAAACCG 220
QY 68 ValMetGluCyAseAAlaCyGlyProGlnProGlnProGlnProGlnProGln 87
Db 221 GTGATGAGTGTGACGCGCTGCGGATGACAGTCAAGTACGAC-CCGCTTACCAGCGT 279
QY 88 ProGlnProLysPro 92

```


Db	280	GCGGCCCTGCTCCA	294
RESULT 13			
LOCUS	AXJ30698	2439 bp	DNA
DEFINITION	Sequence 1207 from Patent WO0194629.	linear	PAT 09-JAN-2002
ACCESSION	AXJ30698		
VERSION	AXJ30698.1	GI:18103676	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G., Horrigan, S., Soppe, D.R. and Weaver, Z. Cancer gene determination and therapeutic screening using signature gene sets		
AUTHORS	Patent: WO 0194629-A 1207 13-DEC-2001;		
JOURNAL	Avalon Pharmaceuticals (US) Location/Qualifiers		
FEATURES	source 1..2439 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"		
BASE COUNT	503 a 758 c 809 g 369 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	5,14e-15	Length:	2439
Score:	262.50	Matches:	58
Percent Similarity:	76.47%	Conservative:	7
Best Local Similarity:	68.24%	Mismatches:	16
Query Match:	44.95%	Indels:	5
DB:	6	Gaps:	1
US-09-696-872-23 (1-109) x AXJ30698 (1-2439)			
Oy	8	LLeuLALALeuaAlaAlaValAlyCySeSerAlaAlaAlayLyBgLySeSerLeuGlYAsp	27
Db	53	CTGTGCACCCTGGCTGGCCCTCGCGCGCGCGAGCACGGGCGAGAGCCGTTGGGC-----	106
Oy	28	CysCySSerAepLeuGLYProGLImetLeuArgLIuGLInGLIuThAsnAlaLeu	47
Db	107	-----TGAGACTGGGGCCCCGAGAAGCTTCGGGAACTCGCAAGAAACCAACGCGCGCTG	160
Oy	48	GINAsPvLaTgASPTryPLeuArgLIuGLInAlaArgLIuethPhelLeuLyAsnThr	67
Db	161	CAGGACGtGCGGGAGCTGGCTGGGGAGAGAGGtCGAGATCATCTTCTTAATAAACG	220
Oy	68	VAlmetGLucYASpADAlCYsgLYProGLInProGLInPolySProGLInProGLInProGLIn	87
Db	221	GTGAATGAGAGTGTGAAGCGCTGGCGGAGTGCAGACAGTACATGACAC-CGGCTTACCAGGCT	279
Oy	88	ProGLInPolySPro 92	
Db	280	GCGGCCCTGCTCCA 294	
RESULT 14			
LOCUS	AXJ30917	2439 bp	DNA
DEFINITION	Sequence 1426 from Patent WO0194629.	linear	PAT 09-JAN-2002
ACCESSION	AXJ30917		
VERSION	AXJ30917.1	GI:18103896	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G., Horrigan, S., Soppe, D.R. and Weaver, Z. Cancer gene determination and therapeutic screening using signature		
AUTHORS	1		
TITLE			

FEATURES	source	location/Qualifiers
JOURNAL	gene sets	Patent: WO 0194629-A 1426 13-DEC-2001;
FEATURES	Avallon Pharmaceuticals (US)	
source	1. .2439	
ORGANISM	/organism="Homo sapiens"	
REFERENCE	/mol_type="genomic DNA"	
AUTHORS	/db_xref="taxon:9606"	
TITLE		
CHARACTERIZATION	Characterization of human and mouse cartilage oligomeric matrix protein	
UNPUBLISHED	Unpublished (1994)	
ORIGINAL SOURCE	Original source text: Homo sapiens cartilage cDNA to mRNA.	
LOCATION/QUALIFIERS	1. .2439	
ORGANISM	/organism="Homo sapiens"	
REFERENCE	/mol_type="mRNA"	
AUTHORS	/db_xref="taxon:9606"	
TITLE	/map="19p13.1"	
CHARACTERIZATION	/cell_type="chondrocyte"	
UNPUBLISHED	/tissue_type="cartilage"	
ORIGINAL SOURCE	/germline	
LOCATION/QUALIFIERS	1. .2439	
ORGANISM	/gene="COMP"	
REFERENCE	26. .2299	
AUTHORS	/gene="COMP"	
TITLE	/standard_name="cartilage oligomeric matrix protein"	

```
/note="putative"
/codon_start=1
/product="matrix protein"
/protein_id="AA57253.1"
/db_xref="GI:602450"
/translation="MVPTACVLLITLALGASGGGSPGSDLGPQMLREIQETNAALQADVWLRQVREITFLKNTVMCDACMQDSVRTGLPSVRLPLHCAFGCFPGVACIQTESGRCGPCPAFTNGSHCTDNECNAPCFPRVCINTSPGRCBACPPGYSGLPTQGVGLAFAPANKOVCTDINECTGQHNCVPNVSCINTRGSPQCPCPGFVGDOASGCCGGAORPCPDGSPSCHEHACVLEHDSRSCVCRVGMANGIILCGRDTLDGFPDEKLRCPEPCQCRKNCVTVPNSGQEDVDRDGDGDCDPAADGCVPRKNCPLVRNPQKRTDERKMGDACNCRSQKNDQKDTDQDSRGDADCDDDIDGKIRMQADNCPKVENSDQKSDSDGIGDACDNCPSKSNPDQADVHDHFGVGDACSDQDQDGDHQSNDNCPVPNSAQEDSDHDGQDADCDDDDDNDGVPDSRNCRLVNPQGEDADRDGVGVCDDFDADKVVLDKIDVCPENAEVTLTDFFAFQTVLDPEDAOIDPMVVVLNQGREGIVOTMNSDPGLAVGYTAFGVDFEGFPHNTVTDVAGFIFGYQDSSFFYVMMKQKQWYMOANPFRVAVAPGIGLAKVKSSTGEGEQLRNALMHTGPTESQVRLMKDPRNCGKDKKSYRMFLQHRPQVGIRVRFYEGPELVADSNVLDITMRGRLGVCFPSQENITWNLRYRCNDTTPEDYETHQLRQA"
sig_peptide
/gene="COMP"
26..85
repeat_region
/note="putative"
290..892
/gene="putative"
/note="putative"
893..1577
/rpc_family="thrombospondin type 2"
/rpc_type=tandem
repeat_region
/note="putative"
893..1577
/rpc_family="thrombospondin type 3"
/rpc_type=tandem
polyA_signal
2420..2425
polyA_site
2439
/gene="COMP"
2439
BASE COUNT 503 a 758 c 809 g 369 t
ORIGIN
Alignment Scores:
Pred. No.: 5,14e-15 Length: 2439
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 9 Gaps: 1
US-09-696-872-23 (1-109) x HUMCOMP (1-2439)
QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
DB 53 CTGCTACCCCTGGCTGGCTCGCGCGCTCCGACAGGCCAGAGCCCGCTGGGC----- 106
QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGluThrAsnAlaAlaLeu 47
DB 107 -----TCAGACCTGGGCCCGACAGATGCTTCGGGAACCTGCAGAAACCAACCGCGCGCTG 160
QY 48 GlnAspValArgAspTyrLeuArgGlnGlnValArgGlnIleThrPheLeuLysAsnThr 67
DB 161 CAGGACGTGGCGGACCTGGCTGGCGACGACGATCAGGAGATCAGCTCCTGAAAAACACG 220
QY 68 ValMetGlnCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGlnProGln 87
DB 221 GTGATGGAGTGTGACGCGTGGCGGATGCAGCAGTACGACAC-CGGCTTACCAAGCGCT 279
QY 88 ProGlnProLysPro 92
DB 280 GCGGCCCTGCTCCA 294
```

Search completed: August 26, 2003, 18:30:17
Job time : 4837.58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 16:04:40 ; Search time 373.301 Seconds
(without alignments) 788.208 Million cell updates/sec

Title: US-09-696-872-23
Perfect score: 584
Sequence: 1 MRWMLGLALAAVCSAAKK.....PKPQKPEPGTGSSEKDEL 109

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+ p2n.model -DB=cgna_1/USPTO.spool/US09696872/runat_26082003_151137_3213/app.query.fasta_1.462
-DB=N-Geneseq_19JUN03 -OPMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plc -NDB=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09696872.@CGN_1.1.0.@runatc_26082003_151137_3213 -NCPUS=6 -ICPU=3
-NO_MMAP -LARGEOUDERY -NEB_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseq_19JUN03:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	584	100.0	369	AA250497	KDEL receptor inh1
2	464	79.5	369	AA250500	KDEL receptor inh1
3	434	74.3	369	AA250493	KDEL receptor inh1
4	428.5	73.4	387	AA250493	KDEL receptor inh1
5	422	72.3	387	AA250501	KDEL receptor inh1
6	414.5	71.0	387	AA250492	KDEL receptor inh1
7	371.5	63.6	369	AA250496	KDEL receptor inh1
8	345.5	59.2	369	AA250495	KDEL receptor inh1
9	337.5	57.8	357	AA250494	KDEL receptor inh1
10	309.5	53.0	312	AA250498	KDEL receptor inh1
11	262.5	44.9	755	AAA47734	Human COMP/TSP-1 c
12	262.5	44.9	925	AAA47735	Human COMP/TSP-2 c
13	262.5	44.9	2439	ABT10887	Human breast cancer
14	262.5	44.9	2439	ABT07751	Breast cancer-asso
15	262.5	44.9	2439	ABT62104	Colon adenocarcino
16	262.5	44.9	2439	ABL62870	Breast cancer rela
17	262.5	44.9	2439	ABL63089	Breast cancer rela
18	262.5	44.9	2439	ACC50112	Breast cancer asso
19	262.5	44.9	2439	ABX76334	Lung cancer-asso
20	262.5	43.9	764	AAK91566	Human CDNA 5'-end
21	262.5	43.9	764	AAK93136	Human CDNA clone r
22	256.5	43.9	1779	AAK94103	Human full-length
23	256.5	43.9	1779	AAK94103	Human nervous syst
24	154.5	26.5	8532	ABA15407	Human nervous syst
25	154.5	26.5	8532	ABA15585	Human breast cancer
26	154.5	26.5	46275	ABT10145	Human full-length
27	153	26.2	4844	AA544633	Human polynucleoti
28	153	26.2	4844	AAK51864	Human polynucleoti
29	149	25.5	144	AAK52848	Human polynucleoti
30	149	25.5	144	AAK52848	VHC-AG-alpha-1 lin
31	145.5	24.9	2916	AAQ67370	VHC-AG-alpha-1 lin
32	145.5	24.9	3074	AAQ67368	Human polynucleoti
33	145.5	24.9	3074	AAQ66455	Human thrombospo
34	145.5	24.9	3074	AAQ66455	Human thrombospo
35	145.5	24.9	3074	AAQ66455	Human thrombospo
36	142.5	24.4	1866	AAQ66455	Human thrombospo
37	138.5	23.7	320	AAQ66455	Human thrombospo
38	138.5	23.7	2820	AAQ66455	Human thrombospo
39	138	23.6	222	AAQ66455	Xenopus thymospo
40	137	23.5	78	AAQ66455	Cassette encoding
41	133	22.8	900	AAQ66455	Nucleotide sequenc
42	127	21.7	12290	AAQ66455	DNA encoding the p
43	126	21.6	96	AAQ66455	E. coli CFT073 gen
44	126	21.6	1172	AAQ66455	Quadruplex DNA SEQ
45	126	21.6	11466	AAQ66455	Sugarcane proline
			24	AAQ66455	pyAc4-Ascl vector

ALIGNMENTS

RESULT 1
AA250497 standard; DNA, 369 BP.
ID AA250497;
AA250497;
XX
XX 23-MAY-2000 (first entry)
XX
XX KDEL receptor inhibitor-6 DNA.
XX
XX KDEL receptor inhibitor; heat shock protein; immune response;
DE oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW infectious disease; allergy; autoimmune disease; ss.
XX
XX Chimeric - Adenovirus E3.
OS Chimeric - Homo sapiens.

```

OS Chimeric - Camelus sp.
XX Key Location/Qualifiers
FT CDS 10..339
FT /tag= a
FT /product= "KDEL receptor inhibitor protein"
FT sig_peptide 10..69
FT /tag= b
FT /note= "Derived from adenovirus E3"
FT mat_peptide 70..336
FT /tag= c
FT /product= "Mature KDEL receptor inhibitor protein"
FT /tag= d
FT /note= "Human COMP pentamerisation domain"
FT misc_feature 97..234
FT /tag= e
FT /note= "Human COMP pentamerisation domain"
FT misc_feature 235..306
FT /tag= e
FT /note= "Camel IgG linker domain"
FT WO200006729-A1.
XX 10-FEB-2000.
XX 28-JUL-1999; 99WO-US17147.
XX 29-JUL-1998; 98US-0124671.
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX Roelman JE, Mayhew M, Hoe MH,
XX WPI; 2000-195296/17.
XX P-PSDB; AAY44963.
XX Inhibitors of the KDEL receptor which comprises an oligomerization
XX domain useful for promoting secretion of proteins which are normally
XX retained within the cell -
XX Disclosure; Fig 6; 87pp; English.
XX PS The patent discloses the use of KDEL receptor inhibitor to promote
XX secretion of proteins that are normally retained within the cell such as
XX heat shock proteins by inhibiting KDEL receptor-mediated return of
XX protein complexes to endoplasmic reticulum. This makes the secreted heat
XX shock proteins more accessible to the immune system and improves immune
XX response to a target antigen. The inhibitor protein comprises several
XX subunits where each subunit comprises an oligomerisation domain and has
XX at its carboxy terminus a region which binds to a KDEL receptor. The
XX target antigen may be associated with diseases including neoplasia such
XX as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX astrocytoma, with defective tumour suppressor genes, oncogenes,
XX infectious diseases, allergy or autoimmune diseases. The present
XX sequence encodes KDEL receptor inhibitor comprising regions encoding a
XX cleavable signal peptide; the oligomerisation domain of human cartilage
XX oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG
XX linker domain and the carboxy-terminal sequence KDEL. The subsequence
XX GDCC is an alteration of rat COMP which provides increased stability via
XX disulphide bonds.
XX SQ Sequence 369 BP, 90 A, 102 C, 120 G, 57 T, 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 6.49e-45 Length: 369
XX Score: 584.00 Matches: 109
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
US-09-696-872-23 (1-109) x AAZ50497 (1-369)
OY 1 MetArgIYrMetIILeuGIIyLeuLeuAAlaLeuAAlaIaValCYeSseRtIAAAlaLysIys 20
|||||

```

```

DB 10 ATGAGTACATGATTATTAAGCTTCTCGCCTTCGCGGACATCTCAGCGCTGCCAAAAA 69
OY 21 GIIYSerSerLeuGIIyGIIyAaPCyCySeSerAaPIeueGIIyProGIImetLeuAArgIueu 40
DB 70 GGAATCAGCCTGGGGGAGAGACTGTTGTCAGACCTGGGCCCCGAGATGCTTGGGAACTG 129
OY 41 GInGIuThrAaenAlaAlaLeuGIIaAaPyAlaAArgAaPTPLeAaGInGIaValAArgIu 60
DB 130 CAGGAaACCAACGGCGCGCTGCAGAGACGTGCGGACGTGCGGCGAGCAAGTCAGGGAG 189
OY 61 ILeThrPheLeuLyAaenThrValImeGIuCyAaPaAlaCYaGIIyProGIImetPro 80
DB 190 ATCAGCTTCTGTAaAAACACCGGTATGAGTGTACACGCGTCGCGGCGCAGCGAGCGG 249
OY 81 LysProGIImetProGIImetProGIImetProGIImetProGIImetProGIImetPro 100
DB 250 AAACCGCAGCCGCGAGCCGAGCCGAGCCGAGAAACCGGAAACCGGAAACCGGAAAGT 309
OY 101 ThrGIYSerSerGIuLyAaPGIuLeu 109
DB 310 ACCGATCATCAGAAaAAGATGAGTTG 336
XX
RESULT 2
AAZ50500
ID AAZ50500 standard; DNA; 369 BP.
XX
AC AAZ50500;
XX
DT 23-MAY-2000 (first entry)
XX
DE KDEL receptor inhibitor-9 DNA.
XX
XX KDEL receptor inhibitor; heat shock protein; immune response;
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease; ss.
XX
OS Chimeric - Adenovirus E3.
OS Chimeric - Homo sapiens.
OS Chimeric - Camelus sp.
XX
XX Key Location/Qualifiers
XX FH 10..339
XX FT /tag= a
XX FT /product= "KDEL receptor inhibitor protein"
XX FT sig_peptide 10..69
XX FT /tag= b
XX FT /note= "Derived from adenovirus E3"
XX FT mat_peptide 70..336
XX FT /tag= c
XX FT /note= "Mature KDEL receptor inhibitor protein"
XX FT misc_feature 97..234
XX FT /tag= d
XX FT /note= "Human TSP4 trimerisation domain"
XX FT misc_feature 235..306
XX FT /tag= e
XX FT /note= "Camel IgG linker domain"
XX
XX WO200006729-A1.
XX 10-FEB-2000.
XX 28-JUL-1999; 99WO-US17147.
XX 29-JUL-1998; 98US-0124671.
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX Roelman JE, Mayhew M, Hoe MH,
XX WPI; 2000-195296/17.
XX P-PSDB; AAY44966.
XX

```

Query Match:

Y: 72.48%

3.14e-31	Length:	365
434.00	Matches:	79
84.40%	Conservative:	13
72.48%	Mismatches:	17
74.32%	Indels:	0
21	Gaps:	0

```

US-09-696-872-23 (1-109) x AAZ50499 (1-369)
OY 1 MetArgTYrMetIleLeuGlyLeuLeuAlaLeuAlaValCysSerAlaAlaValLys 20
DB 10 ATGAGTTCATGATTTTGGCTTCTCGCCCTTGCGGACGTCTCGACGCGCTGCCAAAAA 69
OY 21 GlySerSerLeuGlyGlyAspCysSerAspLeuGlyProGlnMetLeuArgGluLeu 40
DB 70 GGATCCACCTGGGTGGAGACTGTGTGGGAGACGACCAAGCATTGTGACCCAGCTC 129
OY 41 GlnGluThrAspAlaIleLeuGlnAspValArgAspTrpLeuArgGlnValArgGlu 60
DB 130 ACCCTCTTCAACCAAGATCTAGTGAGCTTCGGGACGACATCCGACAGCCAGTGAAGAA 189
OY 61 IleThrPheLeuLysAspThrValMetGluCysAspAlaCysGlyProGlnProGlnPro 80
DB 190 ATGTACTCATCTGGGAACCACTCATGGAGTGTCAAGTGTGCGGTCCGACGCCGACCG 249
OY 81 LysProGlnProGlnProGlnProGlnProLysProGlnProGlnProGlnProGlnGly 100
DB 250 AAACCGGACCGCGACCCGACGCGGAAACCGGACCGGAAACCGGAAACCGGAAAGGT 309
OY 101 ThrGlySerSerGlyLysAspGluLeu 109
DB 310 ACCGATCATCAGAAAAAGATGAGTTG 336
RESULT 4
ID AAZ50493 standard; DNA; 387 BP.
AC AAZ50493;
XX
XX 23-MAY-2000 (first entry)
DT
XX
XX KDEL receptor inhibitor-2 DNA.
XX
XX KDEL receptor inhibitor; heat shock protein; immune response;
KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease; ss.
XX
XX Chimeric - Mus sp.
OS Chimeric - Camelus sp.
OS Chimeric - Rattus sp.
XX
XX Key Location/Qualifiers
FH CDS 10..357
FT /*tag= a
FT /product= "KDEL receptor inhibitor protein"
FT sig_peptide 10..68
FT /*tag= b
FT /note= "Derived from mouse Bsp"
FT mat_peptide 69..354
FT /*tag= c
FT /product= "Mature KDEL receptor inhibitor protein"
FT /*tag= d
FT /*note= "Rat COMP domain"
FT /*tag= e
FT /*note= "Camel IgG linker domain"
FT
FT
XX MO200006729-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17147.
XX
XX 29-JUL-1998; 98US-0124671.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX

```

```

PI Rothman JE, Mayhew M, Hoe MH;
XX
XX WPI; 2000-195296/17.
DR P-PSDB; AAY44959.
XX
XX Inhibitors of the KDEL receptor which comprises an oligomerization
PT domain useful for promoting secretion of proteins which are normally
PT retained within the cell -
XX
XX Disclosure; Fig 2; 87pp; English.
XX
XX The patent discloses the use of KDEL receptor inhibitor to promote
CC secretion of proteins that are normally retained within the cell such as
CC heat shock proteins by inhibiting KDEL receptor-mediated return of
CC protein complexes to endoplasmic reticulum. This makes the secreted heat
CC shock proteins more accessible to the immune system and improves immune
CC response to a target antigen. The inhibitor protein comprises several
CC subunits where each subunit comprises an oligomerisation domain and has
CC at its carboxy terminus a region which binds to a KDEL receptor. The
CC target antigen may be associated with diseases including neoplasia such
CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
CC astrocytoma, with defective tumour suppressor genes, oncogenes,
CC infectious diseases, allergy or autoimmune diseases. The present
CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
CC cleavable signal peptide; the oligomerisation domain from rat cartilage
CC oligomeric matrix protein (COMP); a camel IgG linker domain and the
CC carboxy-terminal sequence KDEU. The subsequence CDCC is an alteration of
CC rat COMP which provides increased stability via disulphide bonds. This
CC is introduced into host cells by suitable vectors.
XX
XX SQ Sequence 387 BP; 93 A; 107 C; 126 G; 61 T; 0 other;
XX
XX Alignment Scores:
XX
XX Pred. No.: 1,05e-30 Length: 387
XX Score: 428.50 Matches: 87
XX Percent Similarity: 80.36% Conservative: 3
XX Best Local Similarity: 77.68% Mismatches: 5
XX Query Match: 73.37% Indels: 17
XX DB: 21 Gaps: 3
XX
US-09-696-872-23 (1-109) x AAZ50499 (1-387)
OY 8 LeuLeuAlaLeuAlaValCysSerAlaAlaValLysGlySerSerLeuGlyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGGTG-----CGGCGCGAGGATCCAGCTGGTGGAGAC 90
OY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAspAlaLeu 47
DB 91 TGTGT-----CCACAGATGCTTCGAGAACTCCAGAGACTTAATGCCGCTG 138
OY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAspThr 67
DB 139 CAAGACGTGAGAGACCTTTCGACAGAGGTCAAGAGATCACTTCTGAAGATACG 198
OY 68 ValMetGluCysAspAlaCysGly-----ProGln 77
DB 199 GTGATGGAATGTGACGCTTGGGAAATGACGCCGACGACCCCGGTAATGATCCGAG 258
OY 78 ProGlnProLysProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlu 97
DB 259 CCGGACCGGAAACCGGACCGGACCGGACCGGAAACCGGAAACCGGAAACCGGAA 318
OY 98 ProGlnGlyThrGlySerSerGlyLysAspGluLeu 109
DB 319 CCGGAGGTACCGGATCATCAGAAAAAGATGAGTTG 354
RESULT 5
AAZ50501
ID AAZ50501 standard; DNA; 444 BP.
XX
XX AAZ50501;
AC
XX
XX 23-MAY-2000 (first entry)
DT

```

XX KDEL receptor inhibitor-10 DNA.
 XX
 DE
 XX KDEL receptor inhibitor; heat shock protein; immune response;
 KM oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 XX infectious disease; allergy; autoimmune disease; ss.
 OS Chimeric - Mus sp.
 OS Chimeric - Rattus sp.
 OS Chimeric - Camelus sp.
 XX
 XX Key Location/Qualifiers
 FT CDS 10..414
 FT /tag= a
 FT /product= "KDEL receptor inhibitor protein"
 FT sig_peptide 10..69
 FT /tag= b
 FT /note= "Derived from mouse Bip"
 FT mat_peptide 70..411
 FT /tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"
 FT misc_feature 154..279
 FT /tag= d
 FT /note= "Rat COMP pentamerisation domain"
 FT misc_feature 310..381
 FT /tag= e
 FT /note= "Camel IgG linker domain"
 XX
 XX WO200006729-A1.
 XX
 XX 10-FEB-2000.
 XX
 XX 28-JUL-1999; 99WO-US17147.
 XX
 XX 29-JUL-1998; 98US-0124671.
 XX
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 XX Rothman JE, Mayhew M, Hoe MH;
 XX WPI; 2000-195296/17.
 XX P-PSDB; AAY44967.
 XX
 XX Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX
 XX Disclosure; Fig 10; 87pp; English.
 XX
 XX The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; a myc-tag; an N-glycosylation sequence; the
 CC oligomerisation domain of rat cartilage oligomeric matrix protein
 CC (COMP); a camel IgG linker domain and the carboxy-terminal sequence
 CC KDEL. The subsequence GDCC is an alteration of rat COMP which provides
 CC increased stability via disulphide bonds.
 XX
 XX Sequence 444 BP; 115 A; 122 C; 134 G; 73 T; 0 other;
 XX
 Alignment Scores: 4.79e-30 Length: 444
 Pred. No.:

Score: 422.00 Matches: 82
 Percent Similarity: 84.85% Conservative: 2
 Best Local Similarity: 82.83% Mismatches: 1
 Query Match: 72.26% Indels: 14
 DB: 21 Gaps: 2
 US-09-696-872-23 (1-109) x AA250501 (1-444)
 QY 21 GlycerSerLeuGlyGlyAAspCysSerAspLeuGlyProGlnMetLeuArgGluLeu 40
 DB 127 GGATCCAGCCTGGGTGGAGACTGTTGT-----CCACGATGCTTCGAGAACTC 174
 QY 41 GlnGluThrAsnAlaAlaLeuGlnAspValArgAspTrpLeuArgGlnValArgGlu 60
 DB 175 CAGAGACTAATGGCGGCTCCAGACGTGAGAGACTCTTGGCAGCAGGTCAAGAG 234
 QY 61 IleThrPheLeuIysAsnThrValMetGluCysAspAlaCysGly-----75
 DB 235 ATCACCTTCCTGAAGAATACCGTGATGGAATGTGACCTTCGGGAATGCAGCCGCAGC 294
 QY 76 -----ProGlnProGlnProGlnProGlnProGlnProGlnProGlnPro 90
 DB 295 ACCCCCGACTAGTCCGACGCCGACCCGAAACCGACGCCGACGCCGACGCCG 354
 QY 91 LysProGlnProLysProGlnProGlnGlyThrGlySerSerGlyLysAspGluLeu 109
 DB 355 AAACCGACCGCGAACCAGAACCGGAAGGTACCGGATCATCAGAAAAAGATGACTTG 411
 RESULT 6
 AA250492
 ID AA250492 standard; DNA; 387 BP.
 XX
 XX AC AA250492;
 XX
 XX DT 23-MAY-2000 (first entry)
 XX
 XX DE KDEL receptor inhibitor-1 DNA.
 XX
 XX KDEL receptor inhibitor; heat shock protein; immune response;
 KM oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 XX infectious disease; allergy; autoimmune disease; ss.
 XX
 XX Chimeric - Mus sp.
 OS Chimeric - Camelus sp.
 OS Chimeric - Rattus sp.
 XX
 XX Key Location/Qualifiers
 FT CDS 10..357
 FT /tag= a
 FT /product= "KDEL receptor inhibitor protein"
 FT sig_peptide 10..68
 FT /tag= b
 FT /note= "Derived from mouse Bip"
 FT mat_peptide 69..354
 FT /tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"
 FT misc_feature 98..222
 FT /tag= d
 FT /note= "Rat COMP domain"
 FT misc_feature 253..324
 FT /tag= e
 FT /note= "Camel IgG linker domain"
 XX
 XX WO200006729-A1.
 XX
 XX 10-FEB-2000.
 XX
 XX 28-JUL-1999; 99WO-US17147.
 XX
 XX 29-JUL-1998; 98US-0124671.
 XX
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 XX

XX Rothman JE, Mayhew M, Hoe MH;
 XX
 DR WPI: 2000-195296/17.
 DR P-PSDB: AA44958.

XX Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -

PS Disclosure: Fig 1, 87pp; English.

XX The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerization domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain from rat cartilage
 CC oligomeric matrix protein; a camel IgG linker domain and the carboxy
 CC -terminal sequence KDEL. This is introduced into host cells by suitable
 CC vectors.

XX Sequence 387 BP, 94 A, 110 C, 125 G, 58 T, 0 other;

Alignment Scores:

Pred. No.: 1.99e-29 Length: 387
 Score: 414.50 Matches: 86
 Percent Similarity: 79.468 Conservative: 3
 Best Local Similarity: 76.798 Mismatches: 6
 Query Match: 70.988 Indels: 17
 DB: 21 Gaps: 3

US-09-696-872-23 (1-109) x AA250492 (1-387)

QY 8 LeuLeuAlaLeuAlaAlaValaCyseAlaAlaAlaLysGlySerSerLeuGlyGlyAsp 27
 DB 40 TTGCTGCTGCTGGCGCGGTG-----CGGCGGAGGATCCAGCTGGGTGA--- 87
 QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnThrAsnAlaLeu 47
 DB 88 -----GACCTAGCCGCCACAGATGCTTCGAGAACTCCAGAGACTAATCGGCGCTG 138
 QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuThrPheLeuLysAsnThr 67
 DB 139 CAAGACGTCGAGAGAGCTTGGCAGCAGGTCACAGAGATCACTTCTGAGAAATACG 198
 QY 68 ValMetGluCysAspAlaCysGly-----ProGln 77
 DB 199 GGTATGGAATGACGCTTGGGAATGACGCCGACGACCCCGGTAAGTCCGAG 258
 QY 78 ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProGlu 97
 DB 259 CCGGAGCCCAAAACCGAGCCGACGCGGACCCGACCCGAAACCGGAAACCGGAA 318
 QY 98 ProGlnGlyThrGlySerSerGluLysAspGluLeu 109
 DB 319 CCGGAAGTACCGGATCATCAAAAAAGATGAGTTG 354

RESULT 7
 ID AA250496 standard; DNA, 369 BP.
 AC AA250496;
 XX
 XX

DT 23-MAY-2000 (first entry)

XX KDEL receptor inhibitor-5 DNA.

XX KDEL receptor inhibitor; heat shock protein; immune response;
 XX oligomerization domain; neoplasia; sarcoma; lymphoma; leukemia;
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KW infectious disease; allergy; autoimmune disease; ss.

OS Chimeric - Mus sp.
 OS Chimeric - Xenopus sp.
 OS Chimeric - Camelus sp.

XX Key Location/Qualifiers

FT CDS 10..339
 FT /tag= a
 FT /product= "KDEL receptor inhibitor protein"

FT sig_peptide 10..68
 FT /tag= b
 FT /note= "Derived from mouse Bip"

FT mat_peptide 69..336
 FT /tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"

FT misc_feature 97..234
 FT /tag= d
 FT /note= "Xenopus TSP4 trimerisation domain"

FT misc_feature 235..306
 FT /tag= e
 FT /note= "Camel IgG linker domain"

XX WO200006729-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17147.

XX 29-JUL-1998; 98US-0124671.

XX (SLOK) SLOAN KETTERING INST CANCER RES.

XX Rothman JE, Mayhew M, Hoe MH;

XX WPI: 2000-195296/17.

XX P-PSDB: AA44958.

XX Inhibitors of the KDEL receptor which comprises an oligomerization

PT domain useful for promoting secretion of proteins which are normally

PT retained within the cell -

XX Disclosure: Fig 5, 87pp; English.

XX The patent discloses the use of KDEL receptor inhibitor to promote

CC secretion of proteins that are normally retained within the cell such as

CC heat shock proteins by inhibiting KDEL receptor-mediated return of

CC protein complexes to endoplasmic reticulum. This makes the secreted heat

CC shock proteins more accessible to the immune system and improves immune

CC response to a target antigen. The inhibitor protein comprises several

CC subunits where each subunit comprises an oligomerisation domain and has

CC at its carboxy terminus a region which binds to a KDEL receptor. The

CC target antigen may be associated with diseases including neoplasia such

CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and

CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of Xenopus
 CC thrombospondin 4 (TSP4) trimerisation domain including an additional
 CC sub-sequence; a camel IgG linker domain and the carboxy-terminal sequence
 CC KDEL. The subsequence GGGC is an alteration of rat cartilage oligomeric
 CC matrix protein which provides increased stability via disulphide bonds.

XX Sequence 369 BP, 94 A, 94 C, 121 G, 60 T, 0 other;

Alignment Scores: 1.58e-25 Length: 369

Pred. No.: 1.58e-25

Score: 371.50 Matches: 68
 Percent Similarity: 79.41% Conservative: 13
 Best Local Similarity: 66.67% Mismatches: 18
 Query Match: 63.61% Indels: 3
 DB: 21 Gaps: 1

US-09-696-872-23 (1-109) x AA250496 (1-369)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysIsgIysSerSerLeuGlyGlyAsp 27
 DB 40 TTGCTGCTGCTGGCGCGGTG-----CGGGCCGAGGATCCAGCTGGGTGAGAC 90
 QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgIuLeuGlnIuThrAsnAlaAlaLeu 47
 DB 91 TGTGTGGTGAAGTCAGACAGACAGTGTATGGCCAGATACCAATGAATCAGATGCTG 150
 QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgIuIleThrPheLeuLysAsnThr 67
 DB 151 GGAGAGCTCCGAGATGTCATGACAGACAGCGTGAAAGACCAATGTTCTTGAGAAACACC 210
 QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
 DB 211 ATTCCAGAAATGCCAGGCTGTGGCCCGACCGCGAAACCGCACGCCGACGCCGAG 270
 QY 88 ProGlnProLysProGlnProLysProGlnProGluProGluGlyThrGlySerSerGluLysAsp 107
 DB 271 CCGCAGCCGAAACCGCACCGCAACCGGAACCGAAGTACCGATCATCAGAAAAAGAT 330
 QY 108 GluLeu 109
 DB 331 GAGTTG 336

RESULT 8

AA250495
 ID AA250495 standard; DNA; 369 BP.

XX AA250495;
 DT 23-MAY-2000 (first entry)
 XX KDEL receptor inhibitor-4 DNA.
 XX KDEL receptor inhibitor; heat shock protein; immune response;
 KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KM infectious disease; allergy; autoimmune disease; ss.
 XX Chimeric - Mus sp.
 OS Chimeric - Camelus sp.
 XX Key Location/Qualifiers
 FH CDS 10..339
 FT /*tag= a
 FT /product= "KDEL receptor inhibitor protein"
 FT sig_peptide 10..68
 FT /*tag= b
 FT /note= "Derived from mouse Bip"
 FT mat_peptide 69..336
 FT /*tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"
 FT 97..234
 FT /*tag= d
 FT /note= "Mouse TSP3 domain"
 FT misc_feature 235..306
 FT /*tag= e
 FT /note= "Camel IgG linker domain"
 XX MO200006729-A1.
 XX 10-FEB-2000.
 XX 28-JUL-1999; 99WO-US171147.

PR 29-JUL-1998; 98US-0124671.
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 PA
 XX
 XX Rothman JE, Mayhew M, Hoe MH;
 XX
 XX WPI; 2000-195296/17.
 DR P-PSDB; AAY44961.

PT Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell
 PS

Disclosure; Fig 4; 87pp; English.

CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of mouse
 CC thrombospondin 3 (TSP3) trimerisation domain including an additional
 CC sub-sequence; a camel IgG linker domain and the carboxy-terminal sequence
 CC KDEL. The subsequence GDC is an alteration of rat cartilage oligomeric
 CC matrix protein which provides increased stability via disulphide bonds.

XX Sequence 369 BP; 87 A; 103 C; 120 G; 59 T; 0 other;

Alignment Scores:

Pred. No.: 3,71e-23 Length: 369
 Score: 345.50 Matches: 64
 Percent Similarity: 76.47% Conservative: 14
 Best Local Similarity: 62.75% Mismatches: 21
 Query Match: 59.16% Indels: 3
 DB: 21 Gaps: 1

US-09-696-872-23 (1-109) x AA250495 (1-369)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysIsgIysSerSerLeuGlyGlyAsp 27
 DB 40 TTGCTGCTGCTGGCGCGGTG-----CGGGCCGAGGATCCAGCTGGGTGAGAC 90
 QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgIuLeuGlnIuThrAsnAlaAlaLeu 47
 DB 91 TGTGTGGTGAAGTCAGACAGACAGTGTATGGCCAGATACCAATGAATCAGATGCTG 150
 QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgIuIleThrPheLeuLysAsnThr 67
 DB 151 GTGAGCTCCGAGATGTCATGACAGACAGCGTGAAAGATGTCATCAGAAAAAGAT 210
 QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
 DB 211 ATTATGAGTGTCAAGTGTGGTCCGACCGCGAAACCGCACGCCGACGCCGAG 270
 QY 88 ProGlnProLysProGlnProLysProGlnProGluProGluGlyThrGlySerSerGluLysAsp 107
 DB 271 CCGCAGCCGAAACCGCACCGCAACCGGAACCGAAGTACCGATCATCAGAAAAAGAT 330
 QY 108 GluLeu 109
 DB 331 GAGTTG 336

RESULT 9

AA250494
 ID AA250494 standard; DNA; 357 BP.

```

XX AA50494;
AC 23-MAY-2000 (first entry)
DT
XX KDEL receptor inhibitor-3 DNA.
XX
DE KDEL receptor inhibitor-3 DNA.
XX
KW KDEL receptor inhibitor; heat shock protein; immune response;
KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW infectious disease; allergy; autoimmune disease; ss.
XX
OS Chimeric - Mus sp.
OS Chimeric - Camelus sp.
XX
FH Key Location/Qualifiers
FT CDS 10..327
FT /tag= a
FT /product= "KDEL receptor inhibitor protein"
FT sig_peptide 10..68
FT /tag= b
FT /note= "Derived from mouse Bip"
FT mat_peptide 69..324
FT /tag= c
FT /product= "Mature KDEL receptor inhibitor protein"
FT misc_feature 97..222
FT /tag= d
FT /note= "Mouse TSP3 domain"
FT misc_feature 223..294
FT /tag= e
FT /note= "Camel Igg linker domain"
XX
XX WO200006729-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17147.
XX
XX 29-JUL-1998; 98US-0124671.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
XX Rothman JE, Mayhew M, Hoe MH;
XX
XX WPI; 2000-195296/17.
XX P-PSDB; AAY44960.
XX
XX Inhibitor of the KDEL receptor which comprises an oligomerisation
XX domain useful for promoting secretion of proteins which are normally
XX retained within the cell -
XX
XX Disclosure; Fig 3; 87pp; English.
XX
XX The patent discloses the use of KDEL receptor inhibitor to promote
XX secretion of proteins that are normally retained within the cell such as
XX heat shock proteins by inhibiting KDEL receptor-mediated return of
XX protein complexes to endoplasmic reticulum. This makes the secreted heat
XX shock proteins more accessible to the immune system and improves immune
XX response to a target antigen. The inhibitor protein comprises several
XX subunits where each subunit comprises an oligomerisation domain and has
XX at its carboxy terminus a region which binds to a KDEL receptor. The
XX target antigen may be associated with diseases including neoplasia such
XX as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX astrocytoma, with defective tumour suppressor genes, oncogenes,
XX infectious diseases, allergy or autoimmune diseases. The present
XX sequence encodes KDEL receptor inhibitor comprising regions encoding a
XX cleavable signal peptide; the oligomerisation domain of mouse
XX thrombospondin 3 (TSP3) trimerisation domain; a camel Igg linker domain
XX and the carboxy-terminal sequence KDEL. The subsequence GPCC is an
XX alteration of rat cartilage oligomeric matrix protein which provides
XX increased stability via disulphide bonds.
XX
XX *Sequence 357 BP; 84 A; 100 C; 114 G; 59 T; 0 other;
XX

```

```

Alignment Scores:
Pred. No.: 1.92e-22 Length: 357
Score: 337.50 Matches: 65
Percent Similarity: 75.49% Conservative: 12
Best Local Similarity: 63.73% Mismatches: 18
Query Match: 57.79% Indels: 7
DB: 21 Gaps: 2

US-09-696-872-23 (1-109) x AA50494 (1-357)
QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysblySGlySerSerLeuGlyGlyAsp 27
DB 40 TYGCTGCTGCTGGGCGCGTG-----CGGCCCGAGGATCCAGCTCGGTGGAGAC 90
QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
DB 91 TGTGTTAAAGCATTTG-----GTCAACCACTCACCCTCTTACACCAATCCTTA 138
QY 48 GlnAspValArgAspTyrPheuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 139 GTGGAGCTTCGGAGACGACATCCGAGACCAAGTGAAGAAATGTCTCATCCGGAACACC 198
QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
DB 199 ATCATGGAGTGTCAAGTGTGGTCCGACCGCAGCCGAAACCGCACCCGACGCCGACG 258
QY 88 ProGlnProLysProGlnProGlnProLysProGlnProGlnGlyThrGlySerSerGlyLysAsp 107
DB 259 CCGGACCCGAAACCGACCGAACCGAACCGAAGATACCGATCATCGAAAAGAT 318
QY 108 GlnLeu 109
DB 319 GAGTTG 324

RESULT 10
AAZ50498
ID AAZ50498 standard; DNA; 312 BP.
XX
XX AAZ50498;
XX
XX 23-MAY-2000 (first entry)
XX
XX KDEL receptor inhibitor-7 DNA.
XX
XX KDEL receptor inhibitor; heat shock protein; immune response;
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease; ss.
XX
XX Chimeric - Adenovirus E3.
XX Chimeric - Homo sapiens.
XX Chimeric - Camelus sp.
XX
XX Key Location/Qualifiers
XX sig_peptide 10..69
XX /tag= a
XX /note= "Derived from adenovirus E3"
XX CDS 10..282
XX /tag= b
XX /note= "KDEL receptor inhibitor protein"
XX mat_peptide 70..279
XX /tag= c
XX /product= "Mature KDEL receptor inhibitor protein"
XX misc_feature 97..177
XX /tag= d
XX /note= "Human PLB pentamerisation domain"
XX misc_feature 178..249
XX /tag= e
XX /note= "Camel Igg linker domain"
XX
XX WO200006729-A1.
XX

```

RESULT 11
ID AAAA7734 standard; DNA; 755 BP.
AC AAAA7734;
XX
DT 08-NOV-2000 (first entry)
DE Human COMP/TSP-1 chimeric protein coding sequence.
KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein; KW thrombospondin; angiogenesis; tumour; treatment; cancer; KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection; KW glaucoma; de.
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers FT 26..754 FT /*tag= a CDS /product= COMP/TSP-1 chimeric protein
XX
PN MO200044908-A2.
XX
PD 03-AUG-2000.
XX
PF 01-FEB-2000; 2000MO-US02482.
XX
PR 01-FEB-1999; 99US-0118053.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PI Lawler JW;
XX
DR WPI; 2000-514823/46. P-PSDB; AAB00040.
XX
PT Nucleic acids encoding chimeric proteins such as cartilage oligomeric PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for PT inhibiting angiogenesis and treating diseases such as cancer
XX
PS Claim 11; Fig 4a-b; 40pp; English.
XX
CC New nucleic acids are described which encode a protein comprising CC the second and third type 1 repeats of human TSP (thrombospondin)-1, CC but not the TGF (transforming growth factor)-beta activation region CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing CC the second and third type-1 repeats and the COMP (cartilage CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1-1 was CC caused inhibition of the growth of tumours in mice models. CC Thus the nucleic acids and proteins may be useful for treating CC angiogenesis related diseases such as cancer (by reducing the rate of CC growth and size of tumours), arthritis, psoriasis, diabetic CC retinopathy, corneal graft rejection, and glaucoma. They may also be CC used for treating human immunodeficiency virus (HIV) infection. CC Anti-angiogenic therapy has little toxicity, does not require the CC therapeutic agent to enter tumour cells or cross the blood-brain CC barrier, controls tumour growth independently of growth of CC tumour cell heterogeneity, and does not induce drug resistance.
XX
SQ Sequence 755 BP; 147 A; 245 C; 236 G; 127 T; 0 other;
Alignment Scores:
Pred. No.: 3..08e-15 Length: 755
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
GB: 21 Gaps: 1

PT	Diagnosing breast cancer in a patient comprises detecting the level of	
PT	gene expression in cell or tissue samples, where a differential gene	
XX	expression is indicative of breast cancer	
PS	Claim 1, SEQ ID NO 1021, 260pp + Sequence Listing, English.	
XX		
CC	The present invention relates to methods of diagnosing breast cancer in a	
CC	patient, which comprise detecting the level of expression in a tissue	
CC	sample of two or more genes selected from those shown in ABT09867-	
CC	ABT1112, where a differential expression of the genes indicates breast	
CC	cancer. The methods are useful in diagnosing, treating, detecting the	
CC	progression, and in monitoring treatment of breast cancer in patients.	
CC	The methods are also useful as a screening tool for agents that modulate	
CC	the onset or progression of breast cancer. The breast cancer genes may be	
CC	used as diagnostic markers for the prediction or identification of the	
CC	malignant state of breast tissue, for confirming the type and progression	
CC	of cancer, and for drug screening and assays. The present sequence is a	
CC	coding sequence of the invention.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub.published_pct_sequences.	
XX		
SO	Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;	
Alignment Scores:		
Pred. No.:	1,14e-14	Length: 2439
Score:	262.50	Matches: 58
Percent Similarity:	76.47%	Conservative: 7
Best Local Similarity:	68.24%	Mismatches: 16
Query Match:	44.95%	Indels: 5
	24	Gaps: 1
US-09-696-872-23 (1-109) x ABT10887 (1-2439)		
QY	8	LeuLeuAlaLeuAlaAlaValCysSerAlaAlaAlaValGlySerSerLeuGlyGlyAsp 27
DB	53	CTGCTACCTGGCGCTGCGCTCGCGGGGTCCGACAGGGCCAGCGCCGTGGGC----- 106
QY	28	CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnIleuThrAsnAlaLeu 47
DB	107	-----TCAGACTGGGCGCCCGCAGATGCTTCGGGAATCTGCAGAAACCAACGGCGGCTG 160
QY	48	GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB	161	CAGGACGTGCGGGGACTGGCTGGCGGACGAGGTACGGGAGATCACGTTCTCGAATAAACG 220
QY	68	ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
DB	221	GTGATGAGGTGTGACGCGCTGCGGAGTACAGATCAGTACGAC-CGGCCTACCCAGCGT 279
QY	88	ProGlnProLysPro 92
DB	280	GCGGCCCCCTGCTCCA 294
RESULT 14		
ID	ABT07751	standard; DNA; 2439 BP.
XX	ABT07751;	
XX	14-NOV-2002 (first entry)	
DE	Breast cancer-associated gene sequence 59.	
XX		
KW	Gene; ds; breast cancer; breast cancer-associated gene sequence;	
XX	drug development; pharmacogenetics; biosensor development.	
XX	unidentified.	
XX		
XX	WO200259377-A2.	
XX		
PD	01-AUG-2002.	

```

XX      24-JAN-2002; 2002MOC-US02242.
PF      XX
PR      24-JAN-2001; 2001US-263965P.
PR      02-FEB-2001; 2001US-265928P.
PR      09-APR-2001; 2001US-282698P.
PR      09-APR-2001; 2001US-0829472.
PR      04-MAY-2001; 2001US-288590P.
PR      29-MAY-2001; 2001US-294443P.
PA      (EOSB-) EOS BIOTECHNOLOGY INC.
PI      Mack DH, Gish KC, Afar D;
DR      WPI; 2002-583738/62.
DR      N-PSDB; ABJ05594.
XX      PT
PT      Detecting a breast cancer-associated transcript in a patient's cell,
PT      useful for diagnosing breast cancer, comprises contacting a biological
PT      sample with a polynucleotide that selectively hybridizes with breast
PT      cancer nucleic acids -
XX      PS
PS      Claim 9; Page 400-401; 410pp; English.

The invention comprises a method of detecting a breast cancer-associated
transcript in a cell from a patient. The method of the invention involves
contacting a biological sample from the patient with a nucleotide that
hybridises to one of the 69 breast cancer-associated gene sequences shown
in the specification. The method of the invention is useful in the
diagnosis or prognosis of breast cancer, and for detecting genes that are
up or down-regulated in breast cancer cells. Genes identified by the
method of the invention can be used in diagnostic purposes and also as
targets for screening for therapeutic compounds that modulate breast
cancer (e.g. hormones or antibodies). Identification of genes that are
over or under expressed in breast cancer can additionally provide high-
resolution, high-sensitivity datasets which can be used in the areas of
diagnostics, therapeutics, drug development, pharmacogenetics, protein
structure and biosensor development. DNA sequences ABT07693 - ABT07761
represent the 69 breast cancer-associated gene sequences of the
invention.

SQ      Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;

Alignment Scores:
Pred. No.:          1,14e-14           Length:         2439
Score:             262.50              Matches:        58
Percent Similarity: 76.47%            Conservative:   7
Best Local Similarity: 68.24%         Mismatches:    16
Query Match:       44,95#             Indels:         5
DB:                24                  Gaps:           1

US-09-696-872-23 (1-109) x ABT07751 (1-2439)

QY      8 LeuLeuAlaLeuAlaAlaValCySerAAlaAlaIySlySGlySerSerGlyIyAsp 27
Dbb|||||:::|||||
Db      53 CTGCACACCTCGGTCTCCTCCGCCGTCGGACAGGCGCACACCCTTGCGC----- 106
QY      28 CyCSeSerAspIengIyProGlnMetLeuArgGluLeuEngInguTrnAsnAlaAlaLeu 47
Dbb|||||-----TCAGACTCGGCCCCGAGAATGCTTCGGGAATCGCAGAAACCAACGCGGCGCTG 166
QY      48 GlNAspValArGaSPrPleuArgngIngnInValArGjuIIEtnrPhelEuLyASnThr 67
Db      161 CAGAGCCTGCGGAGACCTGCTCGCGCAGCAGCAGTCAAGGAGATCATCGTTCTCTAAAAACAAG 220
QY      68 ValMetEclucYAspAlaCYasGlyProGlnInProGlnProIySPProGlnProGln 87
Db      221 GTGATGGAGTGTGACGCTGTCGGAGTAGCAGCAGTCAATGACGAC-CGAGCTAACCCAGCT 279
QY      88 ProGlnProIySPPro 92
Db      280 GCAGCCCTGCTCCA 294

```

XX	RESULT 15
XX	ABL62104
ID	ABL62104 standard; DNA; 2439 BP.
AC	ABL62104;
DT	15-MAY-2002 (first entry)
DE	Colon adenocarcinoma related gene sequence SEQ ID NO:441.
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid; stomach; lung; prostate; pancreas; carcinoma; antitumour; carcinosarcoma; cytostatic; gene therapy; anti-neoplastic; Wilms' tumour; adenocarcinoma; gene; ds.
OS	Homo sapiens.
PN	WO200194629-A2.
PD	13-DEC-2001.
PX	30-MAY-2001; 2001WO-US10838.
PF	05-JUN-2000; 2000US-209473P. PR 05-JUN-2000; 2000US-209531P. PR 18-SEP-2000; 2000US-231133P. PR 18-SEP-2000; 2000US-233617P. PR 20-SEP-2000; 2000US-234009P. PR 20-SEP-2000; 2000US-234034P. PR 20-SEP-2000; 2000US-234052P. PR 22-SEP-2000; 2000US-234509P. PR 22-SEP-2000; 2000US-234567P. PR 25-SEP-2000; 2000US-234923P. PR 25-SEP-2000; 2000US-234924P. PR 25-SEP-2000; 2000US-235077P. PR 25-SEP-2000; 2000US-235082P. PR 25-SEP-2000; 2000US-235134P. PR 25-SEP-2000; 2000US-235280P. PR 26-SEP-2000; 2000US-235637P. PR 26-SEP-2000; 2000US-235638P. PR 27-SEP-2000; 2000US-235711P. PR 27-SEP-2000; 2000US-235720P. PR 27-SEP-2000; 2000US-235840P. PR 27-SEP-2000; 2000US-235863P. PR 28-SEP-2000; 2000US-236028P. PR 28-SEP-2000; 2000US-236032P. PR 28-SEP-2000; 2000US-236033P. PR 28-SEP-2000; 2000US-236034P. PR 28-SEP-2000; 2000US-236109P. PR 28-SEP-2000; 2000US-236111P. PR 29-SEP-2000; 2000US-236842P. PR 29-SEP-2000; 2000US-236891P. PR 02-OCT-2000; 2000US-237172P. PR 02-OCT-2000; 2000US-237173P. PR 02-OCT-2000; 2000US-237278P. PR 02-OCT-2000; 2000US-237294P. PR 02-OCT-2000; 2000US-237295P. PR 02-OCT-2000; 2000US-237316P. PR 03-OCT-2000; 2000US-237425P. PR 03-OCT-2000; 2000US-237598P. PR 03-OCT-2000; 2000US-237604P. PR 03-OCT-2000; 2000US-237606P. PR 03-OCT-2000; 2000US-237608P. PR 01-NOV-2000; 2000US-244867P. PR 01-NOV-2000; 2000US-245084P.
PA	(AVALON PHARM.
PI	Young PE, Augustine M, Carter KC, Ebner R, Endress G, Horrigan S, Soppet DR, Weaver Z; MPI; 2002-188264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX
PS Claim 1, SEQ ID 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205,

Alignment Scores:	
Pred. No.:	1,14e-14
Score:	262.50
Percent Similarity:	76.47%
Best Local Similarity:	68.24%
Query Match:	44.93%
DB:	24
US-09-696-872-23 (1-109) x ABL62104 (1-2439)	
	Length: 2439
	Matches: 58
	Conservative: 7
	Mismatches: 16
	Indels: 5
	Gaps: 1

[illegible]

Search completed: August 26, 2003, 17:06:37
Job time : 376.301 secs

PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,
PI Soppet DR, Weaver Z;
XX
XX
DR *MPI; 2002-188264/24.
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 16:55:35 ; Search time 85.8496 Seconds
(without alignments)
560.408 Million cell updates/sec

Title: US-09-696-872-23
584
Sequence: 1 MRWILGLALAAVCSAAK.....PKPQKPEPEGTGSSEKDEL 109

Scoring table: BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.epool/US09696872/runat_26082003_151138_3255/app_query.fasta_1.462
-DB=Issued_Patents_NA -QEMT=fastcap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=PCT -NOM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USER=US09696872.QCEN 1.1 76 @runat_26082003_151138_3255 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOBJECT -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfltest.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	579	99.1	372	3	US-09-124-671-24
2	459	78.6	372	3	US-09-124-671-30
3	429	73.5	372	3	US-09-124-671-28
4	428.5	73.4	387	3	US-09-124-671-16
5	422	72.3	444	3	US-09-124-671-35
6	414.5	71.0	387	3	US-09-124-671-14
7	371.5	63.6	369	3	US-09-124-671-22
8	345.5	59.2	357	3	US-09-124-671-20
9	337.5	57.8	357	3	US-09-124-671-18
10	304.5	52.1	315	3	US-09-124-671-26
11	145.5	24.9	3074	5	PCT-US93-11725-3
12	138.5	23.7	2820	5	PCT-US93-11725-1

13	122	20.9	521	4	US-09-643-597-324	Sequence 324, App
14	122	20.9	521	4	US-09-480-884A-324	Sequence 324, App
15	122	20.9	521	4	US-09-542-615A-324	Sequence 324, App
16	122	20.9	521	4	US-09-606-421B-324	Sequence 324, App
17	119	20.4	2790	4	US-08-718-661-1	Sequence 1, Appl1
18	118	20.2	17136	4	US-09-453-702B-158	Sequence 158, App
19	109	18.7	816	4	US-09-252-991A-721	Sequence 721, App
20	109	18.7	1605	4	US-09-252-991A-694	Sequence 694, App
21	109	18.7	1674	4	US-09-252-991A-689	Sequence 689, App
22	109	18.7	1677	4	US-09-252-991A-631	Sequence 631, App
23	108.5	18.6	2629	4	US-09-332-184-17	Sequence 17, Appl
24	108	18.5	405	4	US-09-252-991A-6532	Sequence 6532, App
25	108	18.5	546	4	US-09-252-991A-6514	Sequence 6514, App
26	108	18.5	591	4	US-09-252-991A-6500	Sequence 6500, App
27	108	18.5	915	4	US-09-252-991A-6545	Sequence 6545, App
28	108	18.5	2765	4	US-09-620-112D-61	Sequence 61, Appl
29	107	18.3	9758	4	US-09-332-812A-5	Sequence 5, Appl1
30	103	17.6	10348	2	US-08-457-273B-41	Sequence 41, Appl
31	103	17.6	10348	3	US-08-556-419-13	Sequence 13, Appl
32	103	17.6	10348	3	US-09-041-886-14	Sequence 14, Appl
33	103	17.6	10366	1	US-08-246-982A-5	Sequence 5, Appl1
34	103	17.6	10366	1	US-08-453-265-5	Sequence 5, Appl1
35	101	17.3	450	4	US-09-252-991A-2601	Sequence 2601, App
36	101	17.3	588	4	US-09-252-991A-2243	Sequence 2243, App
37	101	17.3	732	4	US-09-252-991A-2440	Sequence 2440, App
38	101	17.3	1083	4	US-09-252-991A-2517	Sequence 2517, App
39	101	17.3	1431	4	US-09-252-991A-2397	Sequence 2397, App
40	100.5	17.2	2369	2	US-08-525-742-1	Sequence 1, Appl1
41	100	17.1	9758	4	US-09-332-812A-5	Sequence 5, Appl1
42	99	17.0	1144	1	US-08-014-943A-1	Sequence 1, Appl1
43	99	17.0	1144	1	US-08-486-421-2	Sequence 2, Appl1
44	99	17.0	1144	1	US-08-470-911-2	Sequence 2, Appl1
45	99	17.0	1144	2	US-08-486-809-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-124-671-24
Sequence 24, Application US/09124671A
Patent No. 6160088
GENERAL INFORMATION:
APPLICANT: Rothman, James
APPLICANT: Mayhew, Mark
TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
FILE REFERENCE: 31488
CURRENT FILING DATE: 1998-07-29
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatsSeq for Windows Version 3.0
SEQ ID NO 24
LENGTH: 372
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chimeric human COMP-KDEL
US-09-124-671-24

Alignment Scores:
Pred. No.: 3,03e-56
Score: 579.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 99.14%
Length: 372
Matches: 108
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-696-872-23 (1-109) x US-09-124-671-24 (1-372)
QY 2 ArgTyrMetIleLeuGlyLeuAlaLeuAlaValCysSerAlaAlaIleValGly 21
DB 16 AGGTACTGATTTTATGAGCTTGCCCTTGGCGGCACTGCGCAAAAAAGGA 75

QY	22	SeqSerLeuGLYGLYAspCySserAspLeuGLYProGlnMetLeuArgGlnLeuGLN	41
Db	76	TCCACCTGGGGGAGACTGTTGTTCAAGACCTGGGCCCCCAGATGGCTTGGGAACTGCAG	135
QY	42	GIuThrAspAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnInValaArgGluIle	61
Db	136	GAAACCAACGGCGCCCTGCAGAGACCGTGGGAGCTGGCTCCGACAGACGTCAAGGAGATC	195
QY	62	ThrPheLeuLYAsnThrValMetGluCYAspAlaCYeGlyProGlnInProGlnProLYs	81
Db	196	ACGTTCTCGAAAAAACAACGGCTGATGAGTGTGACGCGTGGGGCCGACGCCGACCGGAAA	255
QY	82	ProGlnInProGlnInProGlnInProGlnInProLYsProGlnInProLYsProGlnInProLYs	101
Db	256	CCGACGCCGACCCGACGCCGACGCCGAAACCGACCCGAAACCGGAAACCGGAAAGTTACC	315
QY	102	GlySerSerGluLYAspGluLeu	109
Db	316	GGATCATCAGAAAGATGATGATTG	339
RESULT 2			
US-09-124-671-30			
; Sequence 30, Application US/09124671A			
; Patent No. 6160088			
GENERAL INFORMATION:			
; APPLICANT: Rothman, James			
; APPLICANT: Mayhew, Mark			
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS			
; FILE REFERENCE: 31488			
; CURRENT APPLICATION NUMBER: US/09/124,671A			
; CURRENT FILING DATE: 1998-07-29			
; NUMBER OF SEQ ID NOS: 42			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 30			
; LENGTH: 372			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: chimeric human TSP4-KDEL			
US-09-124-671-30			
Alignment Scores:			
Pred. No.: 6,96e-43 Length: 372			
Score: 459.00 Matches: 84			
Percent Similarity: 86.11% Conservative: 9			
Best Local Similarity: 77.78% Mismatches: 15			
Query Match: 78.60% Indels: 0			
DB: 3 Gaps: 0			
US-09-696-872-23 (1-109) x US-09-124-671-30 (1-372)			
QY	2	ArgTrMetCileLeuGlyYleuLeuAlaLeuAlaValCYsSerAlaAlaIleYsleGly	21
Db	16	AGGTCATGATATTAGGCTTGCTCGCCCTTGGCGCATGTGCAGCGCTCCAAAAAAGA	75
QY	22	SerSerLeuGLYGLYAspCYsCySserAspLeuGLYProGlnMetLeuArgGlnLeuGLN	41
Db	76	TCCACCTGGGGGAGACTGTTGTTCAAGACCTGGGCCCCCAGATGGCTTGGGAACTGCAG	135
QY	42	GIuThrAspAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnInValaArgGluIle	61
Db	136	GAAACCAACGGCGCCCTGCAGAGACCGTGGGAGCTGGCTCCGACAGACGTCAAGGAGATC	195
QY	62	ThrPheLeuLYAsnThrValMetGluCYAspAlaCYeGlyProGlnInProGlnInProLYs	81
Db	196	TCATTTTTCGAAAAACCATTAAGTCTGATGATCCAGGCTTGGGCTCCGACGCCGACCGGAAA	255
QY	82	ProGlnInProGlnInProGlnInProGlnInProLYsProGlnInProLYsProGlnInProLYs	101
Db	256	CCGACGCCGACCCGACGCCGACGCCGAAACCGACCCGAAACCGGAAACCGGAAAGTTACC	315
QY	102	GlySerSerGluLYAspGluLeu	109
Db	316	GGATCATCAGAAAGATGATGATTG	339

```

Db      316  GGATCATCAGAAAAAGATGAGTTG  339
                                         |||
RESULT 3
US-09-124-671-28
; Sequence 28, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human TSP3-KDEL
US-09-124-671-28

Alignment Scores:
Pred. No.:      1.52e-39      Length:      372
Score:          429.00      Matches:      78
Percent Similarity: 84.26%      Conservative: 13
Best Local Similarity: 72.22%      Mismatches:  17
Query Match:    73.46%      Indels:      0
Db:             3           Gaps:      0

US-09-696-872-23 (1-109) x US-09-124-671-28 (1-372)
QY      2  ArgTYrMetIleLeuGlyLeuLeuAlaLeuAlaValCysSerAlaAlaValysGly  21
      |||
DB      16  AGGTACAGATTTTAAAGCTTGCTGCCCTTGCGGAGAGTCTGCAGCGCTGCCAAAAAGGA  75
      |||
QY      22  SerSerLeuGlyGlyAAspCysCysSerAspLeuGlyProGluMetLeuArgLuuLeuGln  41
      |||
DB      76  TCCAGCCTGGGTGGAGACTGTTGTGGGAGACGACGACMAAGCATTTGGTCACCCAGCTCACC  135
      |||
QY      42  GluThrAsnaAlaLeuGlnAspValArgAspTrpLeuArgGlnGlnValArgLuuLe  61
      |||
DB      136  CTCTTCACACAGATCTCTGATGGAGCTTGCGGAGCAGACATCCGAGACAGAGTGAAAGAAATG  195
      |||
QY      62  ThrPheLeuLysAsnThrValMetGluCysAspAlaCysGlyProGlnProGlnProLys  81
      ::|
DB      196  TCACCTCACTCCGGAACACCATCATGATGATGTCAGTGTGGGTGCCACGCCGACCGAGAA  255
      |||
QY      82  ProGlnProGlnProGlnProGlnProLysProGlnProLysProGlnProGlnGlyThr  101
      |||
DB      256  CCGCAGCCGCGACCGCCAGCCGAGCGGAAACCGCAGCGGAAACCGGAAAGGTACC  315
      |||
QY      102  GlySerSerGluLysAspGluLeu  109
      |||
DB      316  GGATCATCAGAAAAAGATGAGTTG  339
      |||

RESULT 4
US-09-124-671-16
; Sequence 16, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0

```



```

; SEQ ID NO 16
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric rat COMP=KDEL
US-09-124-671-16

```

Alignment Scores:	
Pred. No.:	1,82e-39
Score:	428.50
Percent Similarity:	80.36%
Best Local Similarity:	77.68%
Query Match:	73.37%
DB:	3
Length:	387
Matches:	87
Conservative:	3
Mismatches:	17
Indels:	5
Gaps:	3

US-09-696-872-23 (1-109) x US-09-124-671-16 (1-387)

Dy 8 LeuLeuAlaLeuAlaValCysSerAlaAlaIylslySGlyseSerrleuglYglYasp 27
||| ||| ||| |||
Dd 40 TTGCTGCTGCGGGCGGTG-----CGGGCCGAGGGATCACGCTGGGTGGAGAC 90
::: ||| ||| |||

Db 91 TGTGT-----CCACAGATGCTTCGAGAACTCCAGGAGACTAATGCCGCCGTG 138

Dy 48 GlnaspValArgaspTrpLeuArgInglInalArgIuIleThrPheLeuysAsnThr 67
 ||||| : : : : :
 Db 139 CAAGACGTAGAGAGACTCTTGGCAGCAGGTCAGAGAGATCACCTTCTGTGAAGATAACG 198

Dy 68 valmetgltcybaspalacgyly-----Profin 77
| | | | |
Db 199 GTGATGGAAITGTGACGCTTGCGGAATTGCAGCCCGCAGCACCCCCGGTACTAGTCCGCAG 258

[illegible]

319 CCGAAGGTACCGGATCATCAGAAAAAGATGAGTTG 354
 Db

RESULT 5
US-09-124-671-35
; Sequence 35, Application US/09124671A

```

; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark

```

FILE REFERENCE: 31488
TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
CURRENT APPLICATION NUMBER: US/09/124,671A
ATTORNEY: JAMES M. HARRIS, JR.

```
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 35  
; LENGTH: 444
```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: NOT AN

```

US-03-124-0/1-33
Alignment Scores:
Pred NO : 1 150-3

Pred. No.:	1.15e-38	length:	444
Score:	422.00	Matches:	82
Percent Similarity:	84.85%	Conservative:	2
Best Local Similarity:	82.83%	Mismatches:	1
Query Match:	72.26%	Indels:	14
DB:	3	Gaps:	2

US-09-696-872-23 (1-109) x US-09-124-671-35 (1-444)

QY 21 GlySerSerLeuIlgIyAspCysCysSerAspLeuIlgIyProGlnMetLeuArgGlnLeu 40
Dp 127 GGAATCCAGCTGGGTGAGACTGTTGT-----CCACAGATGCTTGAGAACTC 174

Dy 4 | G T S C U I I I A B T A L A I A L E U G I N S P V A L T G A S P I I P E U A T G U I G I N G I N V A I T G I U D 80
||| ||| ||| ||| ||| : : : |||
Dd 175 C A G G A G A C T A A T G C G C G C T T G C A A G A C G T T G A G A G A G C T T T T G C G A C A G C A G C T C A A G A G 234

DQ 61 L I E N P I E L U D Y S A S M I N I V A L M E T G C U C Y B A S P A L A C Y S G L /5

D0 235 A T C A C C T T C T G A A G A T A C G G T G A T G G A A T G T G A C C T T G C G G A A T G C A G C C C G C A C G C 294

Db 295 ACCCCCGGTAAGTCCGAGCCGCGAAGCCGAGCCGCGAGCCGCGAGCCG 354

Db 355 AAACCGAGCCGAACCGGAACCGAAGTACCGGATCTCAGAAAAAGATGACTTG 411

US-09-124-671-14
; Sequence 14, Application US/09124671A
; Patent No. 6160088

APPLICANT: Rothman, James
APPLICANT: Mayhew, Mark
APPLICANT: Hoe, Mee

```

; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29

```

```
;; SOFTWARE: FastSeq for windows Version 3.0
;; SEQ ID NO 14
;; LENGTH: 387
;
```

```

;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: chimeric rat COMP-KDEL

```

Alignment Scores:
Pred. No.: 6.61e-38 Length: 387

Percent Similarity:	79.46%	Conservative:	3
Best Local Similarity:	76.79%	Mismatches:	6
Query Match:	70.98%	Indels:	17

US-09-696-8/2-23 (1-109) X US-09-124-6/1-14 (1-387)

Db 40 TTGCTGCTGCGCGCGTG-----CGGCCGAGGATCAAGCCTGGGTGA--- 87

Db 88 -----GACCTAGCCCCACAGATGCTTCGAGAACTCAGGAGACTAATCGCGCCCTG 138

Db 139 CAAGACGTGAGAGCTCTTTCGCAGACAGGTCAGAGAGATCACTTCCGTGAAGAAATACG 198

Db 199 GTGATGGAATGTGACGCTTGGCGAATGCAGCCCGCAGCAGACCCCGCTACTAGTCCCGAG 256

Db 259 CCGCAGCCGAAAACCGCAGCCGCGCAGCCGCGCAGCCGGAACCGCAGCCGAAAACCGGA 318

```
OY 98 ProgluIyThrGIySerSerGIuIySAspGIuIeu 109
DB 319 CCGAAGGTACCGGATCATCAGAAAAAGATGAGTTG 354

RESULT 7
US-09-124-671-22
; Sequence 22, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric Xenopus laevis TSP4-KDEL
US-09-124-671-22

Alignment Scores:
Pred. No.: 3,816-33 Length: 369
Score: 371.50 Matches: 68
Percent Similarity: 79.41% Conservative: 13
Best Local Similarity: 66.67% Mismatches: 18
Query Match: 63.61% Indels: 3
DB: 3 Gaps: 1

US-09-696-872-23 (1-109) x US-09-124-671-22 (1-369)
OY 8 LeuIeuAlaLeuAlaAlaValCySerAlaAlaLysGIySerSerIeuGIyIyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGGTG-----CGGGCCAGGGATCCAGCTGGGTGAGAC 90
OY 28 CyCySerAspIeuGIyProGIuMetIeuArgGIuIeuGIuThrAsnAlaIeu 47
DB 91 TGTGTGTGAGTCAAGACGACGATGATGGCCAGATTAACCAATGATGATGCTG 150
OY 48 GlnAspValArgAspTrpIeuArgGIuIeuValArgGIuIeThrPheIeuIyAsnThr 67
DB 151 GGAGAGCTCCGAGATGTATGAGACAGCGTGAAAGACACATGTTCTTGAAAAACCC 210
OY 68 ValMetGIuCyAspAlaCyGIyProGIuProGIuProGIuProGIuProGIu 87
DB 211 ATTGCAGATGCCAGGCTGTGCGCCGCGAGCCGAAACCCGAGCCGAGCCGAG 270
OY 88 ProGIuProIyProGIuProIyProGIuProGIuProGIuThrGIySerSerGIuIyAsp 107
DB 271 CCGGCGCCGAAACCCGAGCCGAAACCGAAGCGAAGATCCGATCATCAGAAAAAGAT 330
OY 108 GlnIeu 109
DB 331 GAGTTG 336

RESULT 8
US-09-124-671-20
; Sequence 20, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric mouse TSP3-KDEL
US-09-124-671-20

Alignment Scores:
Pred. No.: 2,996-30 Length: 369
Score: 345.50 Matches: 64
Percent Similarity: 76.47% Conservative: 14
Best Local Similarity: 62.75% Mismatches: 21
Query Match: 59.16% Indels: 3
DB: 3 Gaps: 1

US-09-696-872-23 (1-109) x US-09-124-671-20 (1-369)
OY 8 LeuIeuAlaLeuAlaAlaValCySerAlaAlaLysGIySerSerIeuGIyIyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGGTG-----CGGGCCAGGGATCCAGCTGGGTGAGAC 90
OY 28 CyCySerAspIeuGIyProGIuMetIeuArgGIuIeuGIuThrAsnAlaIeu 47
DB 91 TGTGTGTGAGACAGACCAAGCATTTGTCACCCAGCTCACCTCTTCAACAGATCCTA 150
OY 48 GlnAspValArgAspTrpIeuArgGIuIeuValArgGIuIeThrPheIeuIyAsnThr 67
DB 151 GTGAGCTTCGGAGACATCCGAGACCGATGTAAGAAAGTCACTCATCCGAAACCC 210
OY 68 ValMetGIuCyAspAlaCyGIyProGIuProGIuProGIuProGIuProGIu 87
DB 211 ATCATGAGTGTCAAGTGTGTGGTCCGAGCCGAGCCGAAACCCGAGCCGAGCCGAG 270
OY 88 ProGIuProIyProGIuProIyProGIuProGIuProGIuThrGIySerSerGIuIyAsp 107
DB 271 CCGGCGCCGAAACCCGAGCCGAAACCGAAGCGAAGTACCGATCATCAGAAAAAGAT 330
OY 108 GlnIeu 109
DB 331 GAGTTG 336

RESULT 9
US-09-124-671-18
; Sequence 18, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 357
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric mouse TSP3-KDEL
US-09-124-671-18

Alignment Scores:
Pred. No.: 2,236-29 Length: 357
Score: 337.50 Matches: 65
Percent Similarity: 75.49% Conservative: 12
Best Local Similarity: 63.73% Mismatches: 18
Query Match: 57.79% Indels: 7
DB: 3 Gaps: 2
```

Db 859 GCTCCC--CCTGCACCGCCACACAGGCCCA 885

```
RESULT 12
PCT-US93-11725-1
; Sequence 1, Application PC/TUS9311725
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMAN THROMBOSPONDIN-4
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch
; COMPUTER: IBM-compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11725
; FILING DATE: filed herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/985,296
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7005WO
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2820 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Xenopus laevis
; DEVELOPMENTAL STAGE: Stage 45 (germ line)
; PCT-US93-11725-1

Alignment Scores:
Pred. No.: 4.63e-06 Length: 2820
Score: 138.50 Matches: 37
Percent Similarity: 46.30% Conservative: 13
Best Local Similarity: 34.26% Mismatches: 31
Query Match: 23.72% Indels: 27
DB: 5 Gaps: 4

US-09-696-872-23 (1-109) x PCT-US93-11725-1 (1-2820)
QY 23 SerleuGlyGlyAspCyseSerleuGlyProGlnMetLeuArgGluLeuGlnGlu 42
DB 424 AGTGAAGCCGAGCAGACAGAGTACGTCAGACAGACAGATTGATGGCCAGTAACCA 483
QY 43 ThraenAlaLeuGlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuThr 62
DB 484 ATGATTCAGATCCTGGAGAGCTCCGAGATGTCATGAGACAGACAGAGTAAAGACCATG 543
QY 63 PheLeuLysAsnThrValMetGluCysAspAlaCysGlyProGlnProGln--ProLys 81
DB 544 TTCTTGAGAAACACACATTCAGAAATGCCAGGCGCTGCTGATGTCCTGACTTCCCATGG 603
QY 82 ProGlnProGlnProGln-----ProGlnProLysPro----- 92
DB 604 CCAACCAAGTTCCCGACGCGCTACACCACTACACCTCCCAAGCCTCGATGATGATCA 663
QY 93 -----GlnProLysPro 96
DB 664 ACTTATGTTTCAGAGAGTGGGTCATTTGATACAGAGGGCGGCTTCCATATGTGGCGG 723
```

```
QY 97 GlnProGlnGly---ThrGlySer 103
DB 724 TGTCCTGAAGCCTATACAGGCAAC 747

RESULT 13
US-09-643-597-324
; Sequence 324, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hoeken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 369
; SEQ ID NO 324
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-643-597-324

Alignment Scores:
Pred. No.: 3.59e-05 Length: 521
Score: 122.00 Matches: 19
Percent Similarity: 95.83% Conservative: 4
Best Local Similarity: 79.17% Mismatches: 1
Query Match: 20.89% Indels: 0
DB: 4 Gaps: 0

US-09-696-872-23 (1-109) x US-09-643-597-324 (1-521)
QY 76 ProGlnProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLys 95
DB 189 CCTCAGCCTGAGCCGCAACCCCAACCCCAATCACACCCCGCTCAGCCCAACCCCAAG 248
QY 96 ProGlnProGln 99
DB 249 CCTCAGCCCGCAG 260

RESULT 14
US-09-480-884A-324
; Sequence 324, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; APPLICANT: Hoeken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A
; CURRENT FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 324
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-480-884A-324
```


Thi
e
Blank (us
of

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 18:30:26 ; Search time 313.496 Seconds

(without alignments)
781.669 Million cell updates/sec

Title: US-09-696-872-23

Sequence: 1 MRVITGLLALAVCSAAK.....PKPQKPEPGTGSSEKDEL 109

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xjh
-Q=/cgn2_1/USPTO.epool/US09696872/runat_26082003_151139_3337/app_query.fasta.1.462
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPECL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HARPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09696872@cgn_1_1_291@runat_26082003_151139_3337
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSELOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW PUB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	262.5	44.9	755	10	US-09-919-603-4		Sequence 4, Appl1

2	262.5	44.9	925	10	US-09-919-603-6	Sequence 6, Appl1
3	262.5	44.9	2439	10	US-09-954-531-140	Sequence 140, App
4	262.5	44.9	2439	10	US-09-954-531-359	Sequence 359, App
5	262.5	44.9	2439	11	US-09-918-624B-49	Sequence 49, App
6	262.5	44.9	2439	12	US-10-301-822-40	Sequence 40, Appl
7	262.5	44.9	2439	14	US-10-177-293-71	Sequence 71, Appl
8	140	24.0	537	14	US-10-154-971-21	Sequence 21, Appl
9	139	23.8	912	14	US-10-154-971-23	Sequence 23, Appl
10	138.5	23.7	320	9	US-09-815-343-1091	Sequence 1091, Ap
11	126	21.6	1734	14	US-10-156-761-4994	Sequence 4994, Ap
12	126	21.6	9025608	14	US-10-156-761-1	Sequence 1, Appl1
13	122	20.9	521	9	US-09-735-705-324	Sequence 324, App
14	122	20.9	521	10	US-09-850-716A-324	Sequence 324, App
15	122	20.9	521	10	US-09-897-768-324	Sequence 324, App
16	122	20.9	521	12	US-10-117-982-324	Sequence 324, App
17	122	20.9	1082	9	US-09-925-301-441	Sequence 441, App
18	122	20.9	1412	12	US-10-301-822-154	Sequence 154, App
19	122	20.9	2704	14	US-10-106-698-1371	Sequence 1371, Ap
20	122	20.9	3747	13	US-10-071-766-33	Sequence 33, Appl
21	121	20.7	299	9	US-09-294-093B-2226	Sequence 2226, Ap
22	118	20.2	17136	14	US-10-114-170-158	Sequence 158, App
23	116	19.9	5184	10	US-10-260-715-7	Sequence 7, Appl1
24	116	19.9	88421	10	US-09-976-059-1	Sequence 1, Appl1
25	115	19.7	569	10	US-09-881-752A-117	Sequence 117, App
26	115	19.7	958	10	US-09-881-752A-227	Sequence 227, App
27	113.5	19.4	9025608	14	US-10-156-761-1	Sequence 1, Appl1
28	113	19.3	372	13	US-10-079-623-187	Sequence 187, App
29	112.5	19.3	2276	12	US-10-137-870-9	Sequence 9, Appl1
30	112.5	19.3	2276	12	US-10-140-018-9	Sequence 9, Appl1
31	112.5	19.3	2276	12	US-10-140-021-9	Sequence 9, Appl1
32	112.5	19.3	2276	12	US-10-140-274-9	Sequence 9, Appl1
33	112.5	19.3	2276	12	US-10-140-274-9	Sequence 9, Appl1
34	112.5	19.3	2276	12	US-10-140-807-9	Sequence 9, Appl1
35	112.5	19.3	2276	12	US-10-140-922-9	Sequence 9, Appl1
36	112.5	19.3	2276	12	US-10-140-922-9	Sequence 9, Appl1
37	112.5	19.3	2276	12	US-10-140-926-9	Sequence 9, Appl1
38	112.5	19.3	2276	12	US-10-141-698-9	Sequence 9, Appl1
39	112.5	19.3	2276	12	US-10-141-702-9	Sequence 9, Appl1
40	112.5	19.3	2276	12	US-10-141-704-9	Sequence 9, Appl1
41	112.5	19.3	2276	12	US-10-142-421-9	Sequence 9, Appl1
42	112.5	19.3	2276	12	US-10-142-432-9	Sequence 9, Appl1
43	112.5	19.3	2276	12	US-10-142-767-9	Sequence 9, Appl1
44	112.5	19.3	2276	12	US-10-143-033-9	Sequence 9, Appl1
45	112.5	19.3	2276	12	US-10-144-994-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-09-919-603-4
Sequence 4, Application US/09919603
Patent No. US20020137679A1
GENERAL INFORMATION:
APPLICANT: Lawyer, John W.
TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
FILE REFERENCE: 1440.1033-007
CURRENT APPLICATION NUMBER: US/09/919, 603
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: PCT/US00/02482
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/118,053
PRIOR FILING DATE: 1999-02-01
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 755
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion gene
US-09-919-603-4

Alignment Scores:
Pred. No.: 1.27e-22 Length: 755
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 10 Gaps: 1

US-09-696-872-23 (1-109) x US-09-919-603-4 (1-755)

QY 8 LeuLeuAlaLeuAlaAlaValCySeSerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
DB 53 CTGCTACCTCGCTGGCTCGCTCGGACAGGGCCAGACCCCGTGGC----- 106
QY 28 CySeSerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGluThrAsnAlaAlaLeu 47
DB 107 -----TCAGACCTGGGCGCCGACATGCTTCGGAACTGCAGAAACCAACCGCGCTG 160
QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 161 CAGGACGTGGCGGACTGCTCGCGCAGCAGGTCAAGGAGATCAGCTTCCTGAAAAACAG 220
QY 68 ValMetGluCySeAspAlaCySeGlyProGlnProGlnProGlnProGlnProGln 87
DB 221 GTGATGAGGTGACGCGCTGGCGGATGCAGCATGACGAC-CGGCTTACCAGCGT 279
QY 88 ProGlnProLysPro 92
DB 280 GCGGCCCTGCTCCA 294

RESULT 2

US-09-919-603-6
; Sequence 6, Application US/09919603
; Patent No. US20020137679A1
; GENERAL INFORMATION:
; APPLICANT: Lawler, John W.
; TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
; FILE REFERENCE: 1440.1033-007
; CURRENT APPLICATION NUMBER: US/09/919,603
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: PCT/US00/02482
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/118,053
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 925
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion gene
US-09-919-603-6

Alignment Scores:
Pred. No.: 1.61e-22 Length: 925
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 10 Gaps: 1

US-09-696-872-23 (1-109) x US-09-919-603-6 (1-925)

QY 8 LeuLeuAlaLeuAlaAlaValCySeSerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
DB 53 CTGCTACCTCGCTGGCTCGCTCGGACAGGGCCAGACCCCGTGGC----- 106
QY 28 CySeSerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGluThrAsnAlaAlaLeu 47
DB 107 -----TCAGACCTGGGCGCCGACATGCTTCGGAACTGCAGAAACCAACCGCGCTG 160

QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 161 CAGGACGTGGCGGACTGCTCGCGCAGAGGTCAAGGAGATCAGCTTCCTGAAAAACAG 220
QY 68 ValMetGluCySeAspAlaCySeGlyProGlnProGlnProLysProGlnProGlnProGln 87
DB 221 GTGATGAGGTGACGCGCTGGCGGATGCAGCATGACGAC-CGGCTTACCAGCGT 279
QY 88 ProGlnProLysPro 92
DB 280 GCGGCCCTGCTCCA 294

RESULT 3

US-09-954-531-140
; Sequence 140, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cance
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 140
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-140

Alignment Scores:
Pred. No.: 4.97e-22 Length: 2439
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 10 Gaps: 1

US-09-696-872-23 (1-109) x US-09-954-531-140 (1-2439)

QY 8 LeuLeuAlaLeuAlaAlaValCySeSerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
DB 53 CTGCTACCTCGCTGGCTCGCTCGGACAGGGCCAGACCCCGTGGC----- 106
QY 28 CySeSerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGluThrAsnAlaAlaLeu 47
DB 107 -----TCAGACCTGGGCGCCGACATGCTTCGGAACTGCAGAAACCAACCGCGCTG 160
QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 161 CAGGACGTGGCGGACTGCTCGCGCAGCAGGTCAAGGAGATCAGCTTCCTGAAAAACAG 220
QY 68 ValMetGluCySeAspAlaCySeGlyProGlnProGlnProLysProGlnProGlnProGln 87
DB 221 GTGATGAGGTGACGCGCTGGCGGATGCAGCATGACGAC-CGGCTTACCAGCGT 279
QY 88 ProGlnProLysPro 92
DB 280 GCGGCCCTGCTCCA 294
RESULT 4
US-09-954-531-359


```
; Sequence 359, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 359
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-359

Alignment Scores:
Pred. No.: 4.97e-22 Length: 2439
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 10 Gaps: 1

US-09-696-872-23 (1-109) x US-09-954-531-359 (1-2439)
Qy 8 LeuLeuAlaLeuAlaAlaValaCyseSerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
Db 53 CTGCTCACCTCGGTGCTGCTGCGCGCGCTCGGACAGGCGCCGCTGGGC----- 106
Qy 28 CysCysSerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGluThrAsnAlaAlaLeu 47
Db 107 -----TCAGACTGGGCGGCGCCGACAGTCTTCGGGAACTCGAGAAACCAACGCGCGCTG 160
Qy 48 GlnAspValaArgAspTTPLeuArgGlnGlnValaArgGluIleThrPheLeuLysAsnThr 67
Db 161 CAGACGTCGCGGACGTCGCTCGGACAGGTCAGGAGATCAGCTTCGAAAAACACG 220
Qy 68 ValMetGluCysAspAlaCyseGlyProGlnProGlnProGlnProGlnProGln 87
Db 221 GTGATGAGTGTGACGCTGCGGAGTGCAGACGATCAGTACGAC-CGGCTTACCAGCGT 279
Qy 88 ProGlnProLysPro 92
Db 280 GCGGCCCTGCTCCA 294

RESULT 5
US-09-918-624B-49
; Sequence 49, Application US/09918624B
; Publication No. US20030113720A1
; GENERAL INFORMATION:
; APPLICANT: Scheyde, Xiao Min
; APPLICANT: Sonnaese, Thierry
; TITLE OF INVENTION: CDNAS EXPRESSED IN ADIPOCYTE DIFFERENTIATION.
; FILE REFERENCE: PA-0033 US
; CURRENT APPLICATION NUMBER: US/09/918,624B
; PRIOR FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: 60/222,470
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PERL Program
; SEQ ID NO 49
```

```
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GenBank ID No. US20030113720A1 9602449
US-09-918-624B-49

Alignment Scores:
Pred. No.: 4.97e-22 Length: 2439
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 11 Gaps: 1

US-09-696-872-23 (1-109) x US-09-918-624B-49 (1-2439)
Qy 8 LeuLeuAlaLeuAlaAlaValaCyseSerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
Db 53 CTGCTCACCTCGGTGCTGCTGCGCGCGCTCGGACAGGCGCCGCTGGGC----- 106
Qy 28 CysCysSerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGluThrAsnAlaAlaLeu 47
Db 107 -----TCAGACTGGGCGGCGCCGACAGTCTTCGGGAACTCGAGAAACCAACGCGCGCTG 160
Qy 48 GlnAspValaArgAspTTPLeuArgGlnGlnValaArgGluIleThrPheLeuLysAsnThr 67
Db 161 CAGACGTCGCGGACGTCGCTCGGACAGGTCAGGAGATCAGCTTCGAAAAACACG 220
Qy 68 ValMetGluCysAspAlaCyseGlyProGlnProGlnProGlnProGlnProGln 87
Db 221 GTGATGAGTGTGACGCTGCGGAGTGCAGACGATCAGTACGAC-CGGCTTACCAGCGT 279
Qy 88 ProGlnProLysPro 92
Db 280 GCGGCCCTGCTCCA 294

RESULT 6
US-10-301-822-40
; Sequence 40, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamathkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgett, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26) ... (2299)
US-10-301-822-40
```

Alignment Scores:			
Pred. No.:	4,97e-22	Length:	2439
Score:	262.50	Matches:	58
Percent Similarity:	76.47%	Conservative:	7
Best Local Similarity:	68.24%	Mismatches:	16
Query Match:	44.95%	Indels:	5
DB:	12	Gaps:	1

US-09-696-872-23 (1-109) x US-10-301-822-40 (1-2439)			
QY	8	LeuLeuAlaLeuAlaAlaValCySerAlaAlaLysGlySerSerLeuGlyIAsp	27
	53	CTGCTCACCTGCTGCTGCTCGCGCGCTCGGACAGGGCCAGAGCCGTTGGCC-----	106
QY	28	CyCysSerSerLeuGlyProGlnMetLeuArgGlnLeuGlnIuThrIenAlaLeu	47
Db	107	-----TCACACCTGGGCGCCGACAGATGCTTCGGGAACTGACGAGAAACCAACGGCGCTG	160
QY	48	GlnAspValArgAspTrpLeuArgGlnGlnValArgGlnIleThrPheLeuYsaenThr	67
Db	161	CAGACCTGGGCGGACCTGCTGCGGCGACAGTACGAGATCACCTTCTGAAAAACAG	220
QY	68	ValMetGlnCysAspAlaCysGlyProGlnProGlnProLeuProGlnProGlnProGln	87
Db	221	GTGATGAGAGTGATGACGCGTGGGATGACAGTACAGTACAGCAC-CGGCCTACCCAGCGT	279
QY	88	ProGlnProLeuPro 92	
Db	280	GCGGCGCTGCTCCA 294	

RESULT 7	
US-10-177-293-71	Sequence 71, Application US/10177293
Publication No	US20030124128A1
GENERAL INFORMATION:	
APPLICANT:	Lillie, James
APPLICANT:	Glatt, Karen
APPLICANT:	Zhao, Xumei
APPLICANT:	Gannavarpu, Manjula
APPLICANT:	Kannakar, Shubhang
APPLICANT:	Mertens, Maureen
APPLICANT:	Wyer, Vic
APPLICANT:	Wang, Youzhen
APPLICANT:	Xu, Yongyao
APPLICANT:	Hoersch, Sebastian
APPLICANT:	Monahan, John
APPLICANT:	Meyers, Rachel E.
APPLICANT:	Basu Jr., Robert C.
APPLICANT:	Hortobaeyi, Gabriel N.
APPLICANT:	Puztai, Lajos
APPLICANT:	Meric, Funda
APPLICANT:	Sahin, Aysegül
APPLICANT:	Miller, Gordon B.
TITLE OF INVENTION:	COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
TITLE OF INVENTION:	PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE:	MRI-038
CURRENT APPLICATION NUMBER:	US/10/177,293
CURRENT FILING DATE:	2002-06-21
PRIOR APPLICATION NUMBER:	US 60/299,887
PRIOR FILING DATE:	2001-06-21
PRIOR APPLICATION NUMBER:	US 60/301,572
PRIOR FILING DATE:	2001-06-27
PRIOR APPLICATION NUMBER:	US 60/306,501
PRIOR FILING DATE:	2001-07-18
PRIOR APPLICATION NUMBER:	US 60/325,002
PRIOR FILING DATE:	2001-09-25
PRIOR APPLICATION NUMBER:	US 60/362,585
PRIOR FILING DATE:	2002-03-05
PRIOR APPLICATION NUMBER:	US 60/xxx,xxx
PRIOR FILING DATE:	2002-05-14
NUMBER OF SEQ ID NOS:	506
SOFTWARE: FastSeq for Windows Version 4.0	

```

      / SEQ ID NO 71
      / LENGTH: 2439
      / TYPE: DNA
      / ORGANISM: Homo sapiens
      / US-10-177-293-71

Alignment Scores:
Pred. No.:          4,97e-22           Length:         2439
Score:             262.50              Matches:         58
Percent Similarity: 76.47%             Conservative:    7
Best Local Similarity: 68.24%            Mismatches:     16
Query Match:       44.95%               Indels:         5
DB:                14                   Gaps:           1
US-09-696-872-23 (1-109) x US-10-177-293-71 (1-2439)

QY      8 LeuLeuAlaLeuAlaAlaValCysSerxLAlaIaLySgLYSerSerLenGlyGlyAsp 27
        CGTCTCACCTTGCTGGCTCGCCGCCGTCCGACAGAGCCAGACCCTGGTGGGC----- 106
Dy      53 CTGCTCACCTTGCTGGCTCGCCGCCGTCCGACAGAGCTCTTCGGAACTGCAGAAACAACGGCGCTG 160
QY      28 CysCySerAspLeuGlyProGlnMetLeuArgSLuLeuGlnGluThrAanaAlaLeu 47
        107 -----TCAGACTCTGGCCCCGAGAATGCTTCGGAACTGCAGAAACAACGGCGCTG 160
QY      48 GlNapValaIaArgPTripLeuArgGlnGlnValaIaArgGluIethrPheLeuLySaSrThr 67
        161 CAGGAAGTCGCGGAGATGTGCTGCGGACAGCAAGTCAAGGAGATCAAGTTCCGAAAAAACAG 220
Dy      QY      68 ValMetGLuCYASAPALAcysGlyProGlnProGlnProLySProGlnProGln 87
        221 GTGATGAGATGTGAGACGCTGCGGAGATCAAGCAAGTCAGTAAGCCAC-CGGCTTACCAGCT 279
QY      88 ProGlnProLySPro 92
        280 GCGGCCCTGTCTCCA 294
Db      280 GCGGCCCTGTCTCCA 294

RESULT 8
US-10-154-971-21
/ Sequence 21, Application US/10154971
/ Publication No. US2003008074A1
/ GENERAL INFORMATION:
/ APPLICANT: Hamers, Raymond
/ Myldermans, Serge
/ TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -
/ NUMBER OF SEQUENCES: 29
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SPENCER & FRANK
/ STREET: 1100 New York Avenue, N.W., Suite 300 East
/ CITY: Washington
/ STATE: DC
/ COUNTRY: USA
/ ZIP: 20005
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/154,971
FILING DATE: 28-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/945,244
FILING DATE: <Unknown>
APPLICATION NUMBER: PCT/EP96/01725
FILING DATE: 25-APR-1996
APPLICATION NUMBER: EP 95400932.0
FILING DATE: 25-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gollin, Michael A.
REGISTRATION NUMBER: 31,957
REFERENCE/DOCKET NUMBER: GUIDLA 0003

```

```

TELEFAX: 202-414-4040
INFORMATION FOR SEQ ID NO.: 23
SEQUENCE CHARACTERISTICS:
    LENGTH: 912 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURES:
    NAME/KEY: CDS
    LOCATION: 1..909
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-154-971-23

Alignment Scores:
Pred. No.:          1.16e-07      Length:           912
Score:             139.00        Matches:            23
Percent Similarity: 100.00%     Conservative:       0
Best Local Similarity: 100.00%   Mismatches:         0
Query Match:       23.80%       Indels:              0
DB:                Gaps:         0

US-09-696-872-23 (1-109) x US-10-154-971-23 (1-912)

QY 76 ProGlnProGInProlYsPrOGInProlGInProlGInProlGInProlYsPrOGInProlYs 95
Db 412 CCACCAACAAGCCAAATCACCACCAACCAACCAACCGAAGAAGCAAACCAACCAAAA 471
QY 96 ProGluPro 98
Db 472 CCTGAACCC 480

RESULT 10
US-09-815-343-1091
Sequence 1091, Application US/09815343
Patent No. US2001005596A1
GENERAL INFORMATION:
APPLICANT: Meagher, Madeleine
APPLICANT: Xu, Jiangchun
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.504
CURRENT APPLICATION NUMBER: US/09/815.343
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ. ID NOS: 1556
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1091
LENGTH: 320
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(320)
OTHER INFORMATION: n = A,T,C or G
US-09-815-343-1091

Alignment Scores:
Pred. No.:          3.95e-08      Length:           320
Score:             138.50        Matches:            28
Percent Similarity: 53.62%     Conservative:       9
Best Local Similarity: 40.58%   Mismatches:        29
Query Match:       23.72%       Indels:             3
DB:                Gaps:         1

US-09-696-872-23 (1-109) x US-09-815-343-1091 (1-320)

QY 31 AspleUGlYProGlImetceUargLUenuGLngIurThraenaLaalaLeugLnaSPval 50
Db 60 GACTTTAACGGCATGTCTTGSGTCAATTGCACATTAAACCAACTTCCTGGAGAGGTTG 119
QY 51 ArgAsPTprleUrArgLnIngInValArglUlterHrPhelEuLySAenthValmeGlu 70

```

```

US-09-696-872-23 (1-109) x US-09-735-705-324 (1-521)

; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.:          0.185           Length:      9025608
Score:              126.00         Matches:       19
Percent Similarity: 100.00%        Conservative: 5
Best Local Similarity: 79.17%     Mismatches:   0
Query Match:        21.58%         Indels:       0
DB:                  14            Gaps:         0

US-09-696-872-23 (1-109) x US-10-156-761-1 (1-9025608)

OY      76 ProGlnProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLys 95
Db      6065692 CCGCAGCCGGCAGCCGCAGCCGACGCCAGCCGAGCCGAGCCGAGCCGAGCCGACGCCGAG 6065751
OY      96 ProGlnProGln 99
Db      6065752 CCGCACCCCCCAG 6065763

RESULT 13
US-09-735-705-324
; Sequence 324, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Fanger, Neil
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.45SC14
; CURRENT APPLICATION NUMBER: US/09/735,705
; CURRENT FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 419
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 324
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-735-705-324

Alignment Scores:
Pred. No.:          6.74e-06        Length:      521
Score:              122.00         Matches:       19
Percent Similarity: 95.83%        Conservative: 4
Best Local Similarity: 79.17%     Mismatches:   1
Query Match:        20.89%         Indels:       0
DB:                  9             Gaps:         0

```

```

US-09-897-778-324
Alignment Scores:
Pred. No.: 6.74e-06 Length: 521
Score: 122.00 Matches: 19
Percent Similarity: 95.83% Conservative: 4
Best Local Similarity: 79.17% Mismatches: 1
Query Match: 20.89% Indels: 0
DB: 10 Gaps: 0

US-09-696-872-23 (1-109) x US-09-897-778-324 (1-521)
QY 76 ProGluProGluProGlyseProGluInProGluInProGluInProGluInProGluInProGly 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 189 CCTCAGCCTCAGCCCGCAACCCCGCCCAATCACAACCCCGCTCAGCCCGCAACCCCAAG 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 96 ProGluProGlu 99
|||||:|||||:
Db 249 CCTCAGCCCGCAG 260
|||||:|||||:

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 16:41:45 ; Search time 2940.11 Seconds
(without alignments)
901.052 Million cell updates/sec

Title: US-09-696-872-23
Perfect score: 584
Sequence: 1 MRYMIGLALAAVCSAKK.....PKPQKPEPGTGSSEKDEL 109

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgr2_1/USPRO.spool/US09696872/runat_26082003_151137_3336/app.query.fasta.1.462
-DB=EST -QPMT=fastap -SUFFIX=ext -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bites -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09696872.@CGN 1.1.3596.@runat_26082003_151137_3336 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQDRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vtc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	259.5	44.4	518	10 BE809033	BE809033 214373 MA
2	259.5	44.4	604	9 AV612410	AV612410 AV612410
3	249	42.6	496	14 CB712965	CB712965 AMGNNUC:T
4	249	42.6	561	12 BG900312	BG900312 HOA37-1-C
5	248	42.5	337	13 BY350582	BY350582 BY350582
6	248	42.5	359	13 BY335264	BY335264 BY335264
7	246.5	42.2	365	12 BG900764	BG900764 HOA42-1-H
8	246	42.1	350	12 BG897877	BG897877 HOA25-1-G
9	245.5	42.0	328	9 AT751993	AT751993 cml3a10.x
10	245.5	42.0	366	12 BG897326	BG897326 HOA12-1-E
11	245.5	42.0	370	12 BG897214	BG897214 HOA33-1-F
12	245.5	42.0	371	12 BG898406	BG898406 HOA8-1-F1
13	245.5	42.0	372	12 BG899727	BG899727 HOA32-1-A
14	245.5	42.0	373	12 BG896387	BG896387 HOA30-1-E
15	245.5	42.0	374	9 AM068338	AM068338 cm20910.Y
16	245.5	42.0	375	12 BG900556	BG900556 HOA42-1-C
17	245.5	42.0	379	12 BG898027	BG898027 HOA20-1-H
18	245.5	42.0	380	12 BG896332	BG896332 HOA29-1-D
19	245.5	42.0	381	12 BG900343	BG900343 HOA37-1-F
20	245.5	42.0	383	12 BG898382	BG898382 HOA8-1-F4
21	245.5	42.0	384	12 BG896503	BG896503 HOA30-1-C
22	245.5	42.0	384	12 BG899936	BG899936 HOA41-1-E
23	245.5	42.0	384	12 BG900459	BG900459 HOA41-1-C
24	245.5	42.0	385	12 BG897369	BG897369 HOA12-1-B
25	245.5	42.0	385	12 BG897511	BG897511 HOA11-1-A
26	245.5	42.0	385	12 BG898968	BG898968 HOA21-1-C
27	245.5	42.0	385	12 BG900465	BG900465 HOA41-1-C
28	245.5	42.0	386	12 BG898309	BG898309 HOA39-1-A
29	245.5	42.0	386	12 BG898604	BG898604 HOA19-1-G
30	245.5	42.0	386	12 BG899695	BG899695 HOA39-1-D
31	245.5	42.0	388	12 BG896715	BG896715 HOA45-1-C
32	245.5	42.0	390	12 BG897500	BG897500 HOA14-1-G
33	245.5	42.0	390	12 BG898326	BG898326 HOA55-1-C
34	245.5	42.0	390	12 BG898329	BG898329 HOA55-1-D
35	245.5	42.0	391	9 AT751616	AT751616 cml1a01.Y
36	245.5	42.0	392	12 BG898451	BG898451 HOA9-1-D1
37	245.5	42.0	393	9 AM067829	AM067829 cml9f09.x
38	245.5	42.0	393	12 BG898353	BG898353 HOA55-1-D
39	245.5	42.0	394	12 BG900134	BG900134 HOA51-1-D
40	245.5	42.0	395	12 BG899861	BG899861 HOA40-1-C
41	245.5	42.0	397	12 BG898320	BG898320 HOA55-1-C
42	245.5	42.0	399	12 BG897459	BG897459 HOA13-1-H
43	245.5	42.0	399	12 BG899824	BG899824 HOA40-1-F
44	245.5	42.0	401	12 BG896555	BG896555 HOA47-1-D
45	245.5	42.0	401	12 BG901017	BG901017 HOA52-1-G

ALIGNMENTS

RESULT 1
LOCUS BE809033
DEFINITION BE809033 518 bp mRNA linear EST 25-APR-2001
ACCESSION BE809033
VERSION BE809033.1 GI:10240145
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 518)

AUTHORS	Smith, T. P. L., Grose, W. M., Freking, B. A., Roberts, A. J., Stone, R. T., Cessa, E., Wray, J. E., White, J. W., Cho, J., Fahrenkrug, S. C., Bennett, G. L., Heathon, M. P., Laegreid, W. W., Rohrer, G. A., Chitko-Mckown, C. G., Partee, G., Holt, I., Karaymcheva, S., Liang, F., Quackenbush, J., and Keeler, J. W.
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL	Genome Res. 11 (4), 626-630 (2001)
MEDLINE	21180013
PUBMED	11282978
COMMENT	Contact: Smith TPL

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCAGCTACGACG
Plate: 70 row: 1 column: 14
Seq primer: ATTAGTGACACTTAATG.

```

SOURCE
1: 3316
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2BOV"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from testis, thymus,
semitehdonous muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT
85 a 187 c 160 g 86 t
ORIGIN

```

Alignment Scores:	
Pred. No.:	7.59e-15
Score:	259.50
Percent Similarity:	61.74%
Best Local Similarity:	53.91%
Query Match:	44.43%
GB:	10
Length:	518
Matches:	62
Conservative:	9
Mismatches:	21
Indels:	23
Gaps:	3

US-09-696-872-23 (1-109) X BE809033 (1-518)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerSerAlaAlaIysIysb3g5ysSerSerLeuGlyGlyAsp 27
 ||||| |||||::: ::::: |||||
 Db 59 TTGCTCAGCGCTGCGCTGCGCCATCAAGGCCAGGGCAGATGCGCTGGGTGGG--- 115

.....

[illegible]

pb 167 CAGGACGTACGGGACCTACTTGGCGGCACGACAGTCAAGGAGTTACCTTCCTGAAACACACGC 226

68 valMetGluCySAspAlaCysGlyProGlnProGlnProLysProGlnProGlnProGln 87

Db 227 GTGATGAGTGTGACGCGTGCGGATGCAACCGCGGAACCCCAACTGACGGTACGG 286

```

QY      88 PROGLNPROLYSPROGLNPRO----- 94

```

Db 287 CCGCTAGCCAGTGTTCGCCCGGCTTCTGCTTCCCGGAGTGGCTTGTAACGAGACAGCC 346

QY 95 -----LysProGluProGluGly---ThrGlySer 103

347 AACGCGCGCTGCGGACCCCTGCCCCGAAAGTTTCACAGGCAAC 391

2

RESULT 2	LOCUS	DEFINITION	ACCSSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AV612410	AV612410	604 bp mRNA linear EST 28-NOV-2001	AV612410	AV612410	Bos taurus lung fetus Bos taurus clone E11U045D06 5', mRNA sequence.	EST.	Bos taurus (cow)
			AV612410.1	GI:9748080			Bos taurus

REFERENCE
AUTHORS
1 (bases 1 to 604)
Takasuga, A., Hirschune, S., Itoh, R., Jitchazono, A., Suzuki, H., Aso, H.

TITLE Establishment of a high throughput EST sequencing system using

bovine ESTs	30 (33)	E108 (2001)
Wald's acids	30 (33)	E108 (2001)

MEDLINE 21570554
PIIMED 11713338

COMMENT
Contact: Ioshikazu Sugimoto
Animal Genetics Division

Saitakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazunugi@cocoa.ocn.ne.jp

Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

```

1 downloaded | 100%
Source
1 .604
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E11U04SD06"
/tissue_type="lung"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_id="Bos_taurus_lung_fetus"
/note=Vector: pZLI; Site_1: SalI; Site_2: NotI; Poly A

```

BASE COUNT	103 a	203 c	189 g	106 t	3 others
------------	-------	-------	-------	-------	----------

2
2
2
3
3
1
4

3
3
3
3
3
3
.

Pred. No.:	9.07e-15	Length:	604
Score:	259.50	Matches:	62

Percent Similarity:	81.74%	Conservative:	5
Best Local Similarity:	53.91%	Mismatches:	21

DB: 9 Gaps: 3

US-09-696-872-23 (1-109) X AV612410 (1-604)

QY 8 LeuLeuAlaLeuAlaValCysSerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27

D5 70 TTGCTCACGCTGCTGCCCTCGGGCATCAGGCCAGGGCCAGATGCCGCTGGTGGG-- 126

28 cybcyssearbp leuglyf r o g i n m e l l e u a i g i u l e u g n i u i n i a s i w a i d a l e u 4 /

[illegible][illegible]

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

[illegible]

0v a8 prog]nPrOIv8Prog]nPro----- 94


```

Db      298 CCGTAAGCAGTGTGCGCGGCTTCTTCCCGGAGTGGCTTGACCGAGACAGCC 357
Qy      95 -----lysProglinProglinlyl---ThrglySer 103
Db      358 AACGGCGCGGCTCGGACCTCGCCGAGAGTTTCAACAGGCAAC 402

RESULT 3
LOCUS   CB712965
DEFINITION
CB712965          496 bp  mRNA  linear  EST 10-APR-2003
AMNNUC: TRGS2-00012-H5-A trgs2 (10306) Rattus norvegicus cDNA clone
CB712965
CB712965.1  GI:29770113
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 496)
AUTHORS      Angen EST Program.
TITLE        Angen Rat EST Program
JOURNAL      Unpublished
COMMENT      Contact: Dan Fitzpatrick
              Angen, Inc
              One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
              Tel: 805 447-4881
              Plate: 00012 row: h column: 5.
              Location/Qualifiers
FEATURES
    source
        1..496
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /db_xref="taxon:10116"
            /clone="trgs2-00012-H5"
            /tissue_type="muscle"
            /clone_lib="trgs2 (10306)"
            /note="Vector: C6K6P7L; Site_1: SalI; Site_2: NotI; Rat
            gastricentus and soleus muscle"
BASE COUNT   76 a 147 c 129 g 75 t 69 others
ORIGIN
US-09-696-872-23 (1-109) x CB712965 (1-496)

Alignment Scores:
Pred. No.:      6.94e-14      Length:      496
Score:          249.00      Matches:      55
Percent Similarity: 71.26%      Conservative: 7
Best Local Similarity: 63.22%      Mismatches: 21
Query Match:    42.64%      Indels:      4
DB:             14          Gaps:      1

Qy      8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysIsglySerSerLeuGlyIyAsp 27
Db      94 GTGCTCGGCGCTGCGCTTCCGCGGCTACCGGCGCCAGATCCCGCTGGTGGGA--- 150
Qy      28 CysCysSerAspLeuGlyProglinMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
Db      151 -----GACCTAGCCCCCAGACAGATGCTTGAGAACTCCGAGAGACTTAATGCGCGCTG 201
Qy      48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
Db      202 CAAGACGTGAGAGAGCTTTCGACAGCAGGTCAAGAGATCACTTCTTAAGATACG 261
Qy      68 ValMetGluCysAspAlaCysGlyProglinProglinProlysinProglinProglin 87
Db      262 GTGATGAGATGTGACGCTTGGCGAATGACGCCGACACCCCGCGTCTGAGCGTGGG 321
Qy      88 ProglinProlysinProglinPro 94
Db      322 CCAAGTCCGCTCTGCGCACCC 342

RESULT 4

```

```

BG900312
LOCUS   BG900312          561 bp  mRNA  linear  EST 06-NOV-2001
DEFINITION
HOA37-1-C8 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
mRNA sequence.
ACCESSION
BG900312
VERSION
BG900312.1  GI:14310561
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 561)
AUTHORS      Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J.,
              Sathie,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
              Lark,M.W.
              Identification and initial characterization of 5000 expressed
              sequenced tags (ESTs) each from adult human normal and
              osteoarthritic cartilage cDNA libraries
              Osteoarthr. Cartil. 9 (7), 641-653 (2001)
JOURNAL
MEDLINE
21482651
PUBMED
11597177
COMMENT
Contact: Sanjay Kumar
          UW2109
          GlaxoSmithKline
          709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
          Tel: 610-270-7245
          Fax: 610-270-5598
          Email: sanjay_kumar-10@sk.com
          Seq primer: T7.
          Location/Qualifiers
FEATURES
    source
        1..561
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /tissue_type="cartilage"
            /lab_host="E.coli DH10 B"
            /clone_lib="HOA (Human Osteoarthritic Cartilage)"
            /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
            Directional"
BASE COUNT   92 a 212 c 171 g 85 t 1 others
ORIGIN
US-09-696-872-23 (1-109) x BG900312 (1-561)

Alignment Scores:
Pred. No.:      8e-14      Length:      561
Score:          249.00      Matches:      57
Percent Similarity: 73.03%      Conservative: 8
Best Local Similarity: 64.04%      Mismatches: 20
Query Match:    42.64%      Indels:      5
DB:             12          Gaps:      1

Qy      8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysIsglySerSerLeuGlyIyAsp 27
Db      137 CTGTACACCTGCGCTGCGCGGCGCTCGGACAGGCGCAGACCCGTTGGGC----- 190
Qy      28 CysCysSerAspLeuGlyProglinMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
Db      191 -----TCAGACCTGGGCGCCGAGATCTTCGGGAACTGAGAAACCAACGCGGCGCTG 244
Qy      48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
Db      245 CAGGACCTGGGAGAGCTGCTGCGGACAGGTCAAGAGATCACTTCTTAAGATACG 304
Qy      68 ValMetGluCysAspAlaCysGlyProglinProglinProlysinProglinProglin 87
Db      305 GTGATGAGTGTGACGCTGCGGAGATGACAGATCACTAGTACGAC-CCGCTTACCACGCT 363
Qy      88 ProglinProlysinProglinProlysinPro 96
Db      364 CGGCGCCCTGCTCACTGCGCGCCCG 390

RESULT 5

```

BY350582 337 bp mRNA linear EST 12-DEC-2002
 LOCUS BY350582 RIKEN full-length enriched, whole joints Mus musculus cDNA
 DEFINITION clone L830018K02 5', mRNA sequence.
 ACCESSION BY350582
 VERSION BY350582.1 GI:26580070
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 337)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
 Beisel, K.W., Blake, J.A., Brad, D., Brusci, V., Chochia, C., Corbani,
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
 Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ring,
 Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,
 B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempel, C., Sessa,
 M., Shimada, K., Sultana, R., Takenaki, Y., Taylor, M.S., Teasdale,
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
 Watanabe, Y., Wells, C., Wilming, L.G., Wymshaw-Boris, A., Yanagisawa,
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
 Carninci, P., Hayatsu, N., Hirozane-Kikukawa, T., Kono, H., Nakamura,
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Iwami, J., Atzawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
 K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analyses of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 JOURNAL 22354683
 MEDLINE
 PUBMED
 COMMENT
 TITLE
 CONTACT: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Atzawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,
 Murata, M., Nakamura, M., Nomura, K., Nunazaki, R., Ohno, M., Sakai, K.,
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,
 M., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
 Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.
 Tissues were provided by Vassilis Aidinis (Biomedical Sciences
 Research Center 'A1, Fleming' Institute of Immunology 14-16 A1,
 Fleming street 16772 Vari, Greece) whose assistance we gratefully
 acknowledge
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 FEATURES
 source
 1. 337
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="L830018K02"
 /tissue_type="whole joints"
 /clone_id="RIKEN full-length enriched, whole joints"
 BASE COUNT 63 a 108 c 108 g 58 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,56-14 Length: 337
 Score: 248.00 Matches: 58
 Percent Similarity: 65.00% Conservative: 7
 Best Local Similarity: 58.00% Mismatches: 31
 Query Match: 42.47% Indels: 4
 Ds: 13 Gaps: 1
 US-09-696-872-23 (1-109) x BY350582 (1-337)
 QY 8 LeuenaAlaAlaAlaValCySerAlaAlaAlaValysGlySerSerLeuGlyGlyAsp 27
 Db 47 GTCCTGGCCCTGGCTATCTCTGGGCGACAGCCAGCCAGATCCGCTGGTGGGA--- 103
 QY 28 CyCSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGlnTrrAsnAlaAlaLeu 47
 Db 104 -----GACCTGGCCCGCCACACATGCTGGAGAACTTACGAGACGTAATGGCGCGTG 154
 QY 48 GlnAspValArgAspTrrPleuArgGlnGlnValArgGluLeuThrPheLeuLeuAsnThr 67
 Db 155 CAAGACGAGAGAGAGCTGTGCGACAGAGGTCAAGAGATCACCTTCTGAGAAATACG 214
 QY 68 ValMetGluCyAspAlaValCySerGlyProGlnProGlnProGlnProGlnProGln 87
 Db 215 GTGATGGAATGTGATGCTGTGGAGATGAGCCGACGACCTCCAGGCGCTGAGCGTGG 274
 QY 88 ProGlnProLysProGlnProLysProGlnProGlnGlyThrGlySerSerGluLysAsp 107
 Db 275 CCAAGCGCCGCTCGCGACCGCGCTCGTCCCGGCGAGTCTGCTCGAGACAGAT 334
 RESULT 6
 LOCUS BY353264 359 bp mRNA linear EST 11-DEC-2002
 DEFINITION BY353264 RIKEN full-length enriched, synovial fibroblasts Mus
 musculus cDNA clone L130042013 5', mRNA sequence.
 ACCESSION BY353264
 VERSION BY353264.1 GI:26529093
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 359)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
 Beisel, K.W., Blake, J.A., Brad, D., Brusci, V., Chochia, C., Corbani,
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
 Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki

H., Nagashima, T., Numata, K., Okido, T., Pavan, M. J., Perce, G., Pesole, G., Petrovsky, N., Pillai, R., Ponting, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C. A., Secou, M., Shindana, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Matanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yangisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

JOURNAL
MEDLINE
PUBMED
12466851

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Vassilis Aidinis (Biomedical Sciences Research Center, 'A1. Fleming' Institute of Immunology 14-16 A1, Fleming street 1672 Vari, Greece) whose assistance we gratefully acknowledge.

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES
SOURCE

Location/Qualifiers
1. .359
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="U130042033"
/cell_type="synovial fibroblasts"
/clone_lib="RIKEN full-length enriched, synovial fibroblasts"
BASE COUNT 63 a 119 c 117 g 60 t
ORIGIN

Alignment Scores:
Pred. No.: 5.92e-14 Length: 359
Score: 248.00 Matches: 58
Percent Similarity: 65.66% Conservative: 7
Best Local Similarity: 58.59% Mismatches: 30

Query Match: 42.47% Indels: 4
DB: 13 Gaps: 1
US-09-696-872-23 (1-109) x BY355264 (1-359)

Qy 8 leuleuAlaAlaAlaAlaValCySserAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
Db 47 GTGCTCCCTGGCTATCTCGCGGCGACAGCGCAGGCGCAGATCCCGCTGGGTGGA--- 103
Qy 28 CySCysSerSerLeuGlyProGlnMetLeuArgLysLysLysLysLysLysLysLys 47
Db 104 -----CACCTGGCCCAAGATCTCGAGAACTTACGAGACTTAATGCGCGCTG 154
Qy 48 GlnAspValArgAspTyrPLeuArgGlnGlnValArgLysLysLysLysLysLysLys 67
Db 155 CAGACCTGAGAGAGCTCTTCCAGACAGCTCAGAGATCACTTCTCGAAGAAATGCG 214
Qy 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
Db 215 GTGATGAAATGATGATGCTTCCGGAATGACCGCCAGCACTCAGGCTGAGGCTCGG 274
Qy 88 ProGlnProLysProGlnProLysProGlnProGlnGlyThrGlySerSerGlyLys 106
Db 275 CCAATGCGCTCTGCGACCGCGCTCTCTCCCGCGCTGATGCTCTCGAGAAA 331

RESULT 7

LOCUS BG900764 365 bp mRNA linear EST 06-NOV-2001

DEFINITION HNA42-1-H5 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.

ACCESSION BG900764
VERSION BG900764.1 GI:14311013

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 365)
Kumar, S., Connor, U.R., Dodds, R.A., Halsey, W., Van Horn, M., Mao, J.,
Lark, M.W.

Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthritic cartilage cDNA libraries

Osteoarthr. Cartil. 9 (7), 641-653 (2001)

JOURNAL MEDLINE 21482651
PUBMED 11597177

COMMENT Contact: Sanjay Kumar
UN2109

GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay.kumar-1@sk.com
Seq primer: T7.

FEATURES
SOURCE

Location/Qualifiers
1. .365
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/clone_lib="HOA (Human Osteoarthritic Cartilage)"
/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 65 a 123 c 121 g 56 t
ORIGIN

Alignment Scores:
Pred. No.: 8.33e-14 Length: 365
Score: 246.50 Matches: 55
Percent Similarity: 75.61% Conservative: 7
Best Local Similarity: 67.07% Mismatches: 9

DB:	4	9	1	Gaps:
US-09-696-872-23	(1-109)	x AT751993	(1-328)	
OY	8	LeuLeuAlaLeuAlaValCysSerAlaAlaValSerSerLeuGlyCysAsp	27	
Db	56	CTGCTCACCCCTGGCTGGCTGGCCGCGCCGACAGGACAGAGCCGCTGGGCG-----	109	
OY	28	CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGlnIleThrAsnAlaAlaLeu	47	
Db	110	-----TCAGACCTGGGCGCCGAGATGCTTGGGAACTGCAGAAACCAACGCGCGCTG	163	
OY	48	GlnAspValaArgAspTyrLeuArgGlnGlnValaArgGluIleThrPheLeuIysAsnThr	67	
Db	164	CAGAGCTGCCGGAGCTGCTGGCGAGAGGTCAAGGATCAGCTTCTTGAATAAACG	223	
OY	68	ValMetGlnCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln	87	
Db	224	GTGATGAGATGTGACGCGTGGCGGATGACAGTCAATGACGAC-CGACTTACCAGCGT	282	
OY	88	ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln	92	
Db	283	GCGGCGCTGCTCCA	297	
RESULT 10				
LOCUS	BG897326	366 bp	mRNA	linear
DEFINITION	HOA12-1-E3 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.			EST 06-NOV-2001
ACCESSION	BG897326			
VERSION	BG897326.1			
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 366)			
AUTHORS	Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathie,G., Mu,J.P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lack,M.W.			
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries			
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)			
MEDLINE	21482651			
PUBMED	11597177			
COMMENT	Contact: Sanjay Kumar UW2109 GlaxoSmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay_kumar-1@sk.com Seq primer: T7.			
FEATURES	Location/Qualifiers			
source	1..366			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/tissue_type="cartilage"			
	/lab_host="E.coli DH10 B"			
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"			
	/note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI; Directional"			
BASE COUNT	56 a 134 c 117 g 59 t			
ORIGIN				
Alignment Scores:				
Pred. No.:	1.04e-13	Length:	366	
Score:	245.50	Matches:	56	
Percent Similarity:	75.29%	Conservative:	8	
Best Local Similarity:	65.88%	Mismatch:	17	
Query Match:	42.04%	Indels:	5	

DB:		12	Gaps:	1
UN-09-696-872-23	(1-109) x BG897326	(1-366)		
OY	8	LeuDeuAlaIeuAlaIaValCySSerAlaAlaIylSglYSerSerLeuGIyGlyAsp	27	
Db	96	CTGTCAACCGTGGCTGCCTCCGCGGCCGTCAGAGCCAGAGCCCGGTGGC-----	149	
OY	28	CysCySerSerApleuGIyProGIInMetLeuArGIleuGIInGIuThrAsnAlaAlaIeu	47	
Db	150	-----TCAGACCCTGGGCCCCGACAATCTTCGGAACTGCAGGAACCAACGCCGCGCTG	203	
OY	48	GlInAPyAlaArgASPTrPleuArGIInGIInValaArgIuIeThrPheLeuIyASnThr	67	
Db	204	CAGAACGTGCCGGAGCTGCTGCCGACAGGTCAAGGAGATCAAGTTCTGAAAAACACG	263	
OY	68	ValMetGIuCysAspAlaCYsgIyProGIInProGIInProGIInProGIInProGIIn	87	
Db	264	GTGATGAGTGCTGACGCGTGGGATGCAGACATCACAGCCAC-CGAGCTACCCAGCGT	322	
OY	88	ProGIInPolyPro 92		
Db	323	GCGGCCCCCTGCTCCA 337		
RESULT 11				
LOCUS	BG897214	370 bp.	mRNA	linear EST 06-NOV-2001
DEFINITION	H0A33-1-F1 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,			
ACCESSION	BG897214			
KEYWORDS	mRNA sequence.			
SOURCE	BG897214.1 GI:14307455			
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 370)			
JOURNAL	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,			
MEDLINE	Sathe,G., Mut,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and			
PUBMED	Lark,M.W.			
COMMENT	Identification and initial characterization of 5000 expressed			
	sequenced tags (ESTs) each from adult human normal and			
	osteoarthritis cartilage CDNA libraries			
	Osteoarthr. Cartil. 9 (7), 641-653 (2001)			
	21482651			
	11597177			
	Contact: Sanjay Kumar			
	UM2109			
	GlaxoSmithKline			
	709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA			
	Tel: 610-270-7245			
	Fax: 610-270-5598			
	Email: sanjay_kumar-legsk.com			
	Seq primer: T7.			
FEATURES	Location/Qualifiers			
source	1..370			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/tissue_type="cartilage"			
	/lab_host="E.coli DH10 B"			
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"			
	/note="Vector: pSPORT I, Site_1: SalI, Site_2: NotI;			
	Directional"			
BASE COUNT	67 a 127 c 120 g 56 t			
ORIGIN				
Alignment Scores:				
Pred. No.:	1.05e-13	Length:	370	
Score:	245.50	Matches:	55	
Percent Similarity:	74.3%	Conservative:	6	
Best Local Similarity:	67.0%	Mismatch:	10	
Query Match:	42.04%	Indels:	11	

DB:	12	Gaps:	2
US-09-696-872-23 (1-109) x BG897214 (1-370)			
OY	8	LeuenaalaleuaalaalValCyseera1aa1ataylsygyserSerleuglyGlyasp	27
Db	137	CTGCCTACCCCTGGCTGCCCTCGGCCGTCGGACAGAGGCCAAGGCCGTGGG-----	190
OY	28	CysCysSerapleuglyProglInMedleuarvgIuleugInglnlThAsna1aalau	47
Db	191	-----TCAGACTTGGGCCCGCAGATGCTTGCGGAACYGACGAAACCAACGCGCGCTG	244
OY	48	GlNapVlaLgaaptTrypleuarvgInglnnValargSuileThPhleulyasnthr	67
Db	245	CGAGACGCGCGAGACTCTCGCGGACGAGCGACGAGATCACCTTCTGAAAAAACG	304
OY	68	ValMetGluCyaspAlaCyeglyProglIn-----ProglInPro	80
Db	305	GTGATGAGACTGTGACGCTCGCGGATGACAGCATGACGACCGGCGCTACCAGCTG	364
OY	81	LysPro 82	
Db	365	CGGCC 370	
RESULT 12			
BG898406		371 bp	mRNA linear EST 06-NOV-2001
LOCUS	HOM8-1-F10 HOM (Human Osteoarthritic Cartilage)		Homo sapiens cDNA,
DEFINITION			mRNA sequence.
ACCESSION	BG898406		
VERSION	BG898406.1	GI:14308655	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eumaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
REFERENCE	1 (bases 1 to 371)		
AUTHORS	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Lach,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.		
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritis cartilage cDNA libraries		
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)		
MEDLINE	21482651		
PUBMED	11597177		
COMMENT	Contact: Sanjay Kumar UM2109		
	GlxosmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel.: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@sk.com		
FEATURES	Seq primer: T7. Location/Qualifiers		
source	1..371		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/tissue_type="cartilage"		
	/lab_host="E.coli DH10 B"		
	/clone_lib="HOM (Human Osteoarthritic Cartilage)"		
	/note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI; Directional"		
BASE COUNT	65 a 126 c 118 g 62 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,05e-13	Length:	371
Score:	245.50	Matches:	56
Percent Similarity:	75.2%	Conservative:	8
Best Local Similarity:	65.8%	Mismatches:	17
Query Match:	42.04%	Indels:	5

DB:	12	Gaps:	1
US-09-696-872-23	(1-109) x BG898406	(1-371)	
OY	8	LeuDeuAlaIeuAlaAValaCySerAlaIalalySgIySerSerLeuGlyAsp	27
Db	105	CTGGCTACACCCTGGGTGCCCTCGGGCCGTCGGAGCAGAGGCCAGAGCCCGTGGGC-----	158
OY	28	CysCySerApPLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaIleu	47
Db	159	-----TCAGACCTGGGGCCCCAGATGCTTCGGAACTGCAGGAACCAACGCGCGCTG	212
OY	48	GlnAspValaAgasApPTpLeuArgGlnAlaIArgGluIeThrPpeLeuLysAsnThr	67
Db	213	CAGAACCGTGGGGAGCTGCTGCGGACAGAGGTACAGGAGATCACATTCTTAATAAACACG	272
OY	68	ValMetGluCyAsAPalaCySgIyProGlnInProGlnProGlnProGlnInProGln	87
Db	273	GTGATGAGTGtGTGACGGCTGGGATGACAGACAGTCAGTACGCAC-CGGCCTACCAAGCGT	331
OY	88	ProGlnProLysPro	92
Db	332	GCGGCCCTGCTCCA	346
RESULT 13			
BG899727		372 bp	mRNA linear EST 06-NOV-2001
LOCUS	HOA32-1-A8 HOA (Human Osteoarthritic Cartilage)		Homo sapiens cDNA,
DEFINITION	mRNA sequence.		
ACCESSION	BG899727		
VERSION	BG899727.1	GI:14309992	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 372)		
AUTHORS	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sache,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.		
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage CDNA libraries		
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)		
MEDLINE	21482651		
PUBMED	11597177		
COMMENT	Contact: Sanjay Kumar IM2109 GlaxoSmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-l@gsk.com Seq primer: T7. Location/Qualifiers		
FEATURES	source	1..372	
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/cissue_type="cartilage"		
	/lab_host="E.coli DH10 B"		
	/clone_id="HOA (Human Osteoarthritic Cartilage)"		
	/note="Vector: pSPORT I, Site_1: SalI; Site_2: NotI; Directional"		
BASE COUNT	68 a	130 c	117 g 57 t
ORIGIN			
Alignment Scores:	1.06e-13	Length:	372
Pred. No.:	245.50	Matches:	55
Score:	74.39%	Conservative:	6
Percent Similarity:	67.07%	Mismatches:	10
Best Local Similarity:	42.04%	Indels:	11
Query Match:			

Percent Similarity: 74.12%
Best Local Similarity: 65.88%
Query Match: 42.04%
DB: 9
Conservative: 7
Mismatches: 18
Indels: 5
Gaps: 1

US-09-696-872-23 (1-109) x AW068338 (1-374)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysGlySerSerLeuGlyGlyAsp 27
DB 51 CTGCTCACCTGGCTGGCCCTCGCGCGGACAGGCGCAGAGCCGTTGGC----- 104
QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
DB 105 -----TCAGACCTGGCGCGCGCAGATGCTTCGCGAACTGCAGAAACCAACGCGCGCTG 158
QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 159 CAGGACGTGCGGAGCTGCTGCGGACAGGTCAAGGAGATCAGTCTGAAAAACACG 218
QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
DB 219 GTGATGAGTGTGACCGCGGAGTGCAGACGTCAGT-ATGCACCGGCTTACCAAGCGT 277
QY 88 ProGlnProLysPro 92
DB 278 GCGGCCCTGCTCCA 292

Search completed: August 26, 2003, 19:21:12
Job time : 2942.11 secs

PF	28-JUL-1999;	99WO-US17147.
XX		
PR	29-JUL-1998;	98US-0124671.
XX		
PA	(SLOC) SLOCAN KETTERING INST CANCER RES.	
XX		
PI	Rochman JE, Mayhew M, Hoe MH;	
XX		
DR	WPI; 2000-195296/17.	
PT	N-PSDB; AAZ50497.	
XX		
PS	Disclosure; Fig 6; 87pp; English.	
XX		
CC	The patent discloses the use of KDEL receptor inhibitor to promote	
CC	secretion of proteins that are normally retained within the cell such as	
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of	
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat	
CC	shock proteins more accessible to the immune system and improves immune	
CC	response to a target antigen. The inhibitor protein comprises several	
CC	subunits where each subunit comprises an oligomerisation domain and has	
CC	at its carboxy terminus a region which binds to a KDEL receptor. The	
CC	target antigen may be associated with diseases including neoplasia such	
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and	
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,	
CC	infectious diseases, allergy or autoimmune diseases. The present	
CC	sequence is KDEL receptor inhibitor comprising regions encoding a	
CC	cleavable signal peptide; the oligomerisation domain of human cartilage	
CC	oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG	
CC	linker domain and the carboxy-terminal sequence KDEL. The subsequence	
CC	GDCC is an alteration of rat COMP which provides increased stability via	
CC	disulphide bonds.	
XX		
SO	Sequence 109 AA;	
QY	Query Match 100.0%; Score 584; DB 21; Length 109;	
	Best Local Similarity 100.0%; Pred. No. 7.1e-52;	
	Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MRWIMIGLALAAVCSAAKKGSSIGDCCSDLGPMLELOETNALDVRDWMRQVRE 60	
DB	1 MRYWILGLALAAVCSAAKKGSSIGDCCSDLGPMLELOETNALDVRDWMRQVRE 60	
QY	61 ITPLKNTIMECDACGPPQPPKPPQPPQPPKPPKPPPEEGTSSSEKDEL 109	
DB	61 ITPLKNTIMECDACGPPQPPKPPQPPQPPKPPKPPPEEGTSSSEKDEL 109	
RESULT 2		
AC	AAAY4966	
XX	AAAY4966 standard; Protein; 109 AA.	
XX	AAAY4966;	
DT	23-MAY-2000 (first entry)	
DE	KDEL receptor inhibitor protein-9.	
XX		
KW	KDEL receptor inhibitor; heat shock protein; immune response;	
KW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;	
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;	
KW	infectious disease; allergy; autoimmune disease.	
XX		
OS	Chimeric - Adenovirus B3.	
OS	Chimeric - Homo sapiens.	
OS	Chimeric - Camelus sp.	
XX		
EH	Key Location/Qualifiers	
FT	Peptide 1..20	
FT	/label= Signal peptide	

PT	Domain	30..75	/note= "Derived from adenovirus E3"
FT	Domain	/note= "Human TSP4 trimerisation domain"	
FT	Domain	76..99	
FT	Domain	/note= "Camel IgG linker domain"	
PN	WO200006729-A1.		
XX			
PD	10-FEB-2000.		
XX			
PF	28-JUL-1999;	99WO-US17147.	
XX			
PR	29-JUL-1998;	98US-0124671.	
XX			
PA	(SLOK) SLOAN KETTERING INST CANCER RES.		
PI	Rochman JE, Mayhew M, Hoe MH;		
XX			
DR	WPI; 2000-195296/17.		
DR	N-PSDB; AAZ50500.		
PT	Inhibitors of the KDEL receptor which comprises an oligomerization domain useful for promoting secretion of proteins which are normally retained within the cell		
XX			
PS	Disclosure; Fig 9; 87pp; English.		
XX			
CC	The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several subunits where each subunit comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and infectious diseases, allergy or autoimmune diseases. The present invention is a KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; the oligomerisation domain of human thrombospondin 4 (TSP4) trimerisation domain; a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric matrix protein which provides increased stability via disulphide bonds.		
CC			
XX			
XX	Sequence 109 AA;		
CC			
CC	Query Match 79.5%; Score 464; DB 21; Length 109;		
CC	Best Local Similarity 78.0%; Pred. No. 1.2e-35;		
CC	Matches 85; Conservative 9; Mismatches 15; Indels 0; Gaps 0;		
QY	1 MRYWIIIGLILAAVCAVSAKKGSSLGSDCCSDLPQMLRELQETNAALQDYVDMRLQVRE 60		
Db	1 MRYWIIIGLILAAVCAVSAKKGSSLGSDCCSDLPQMLRELQETNAALQDYVDMRLQVRE 60		
QY	61 ITFLAKTVMKCDACGPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 109		
Db	61 ITFLAKTVMKCDACGPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 109		
RESULT 3			
AAV44965			
ID	AAV44965 standard; Protein; 109 AA.		
AC	AAV44965;		
XX			
DT	23-MAY-2000 (first entry)		
XX			
DE	KDEL receptor inhibitor protein-8.		
XX			
XX	KDEL receptor inhibitor; heat shock protein; immune response; oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;		
XX			

KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KM infectious disease; allergy; autoimmune disease.
 XX
 OS Chimeric - Adenovirus E3.
 OS Chimeric - Homo sapiens.
 OS Chimeric - Camelus sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= Signal peptide
 FT /note= "Derived from adenovirus E3"
 FT Domain 30..75
 FT /note= "Human TSP 3 trimerisation domain"
 FT Domain 76..99
 FT /note= "Camel IgG linker domain"
 XX
 PN WO200006729-A1.
 XX
 XX 10-FEB-2000.
 XX
 XX 28-JUL-1999; 99WO-US17147.
 XX PF
 XX 29-JUL-1998; 98US-0124671.
 XX
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 PI Rothman JE, Mayhew M, Hoe MH;
 XX
 XX WPI; 2000-195296/17.
 DR N-PSDB; AAZ50499.
 XX
 XX Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX
 XX Disclosure; Fig 8; 87pp; English.
 XX
 CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence is KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of human
 CC thymospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an
 CC alteration of rat cartilage oligomeric matrix protein which provides
 CC increased stability via disulphide bonds.
 XX
 XX Sequence 109 AA;
 SQ
 Query Match 74.3%; Score 434; DB 21; Length 109;
 Best Local Similarity 72.5%; Pred. No. 1.4e-36;
 Matches 79; Conservative 13; Mismatches 17; Indels 0; Gaps 0;
 QY 1 MRWIIIGLLAAVCSAKKSSIGDCCSDLGOMRLQETNALQDVRLKQVRE 60
 DB 1 MRWIIIGLLAAVCSAKKSSIGDCCSDLGOMRLQETNALQDVRLKQVRE 60
 QY 61 ITFLKNTVMECDACGPOPKPOPKPOPKPOPKPEEGTSSSEKDEL 109
 DB 61 MSJLRNTIMECGVCGPOPKPOPKPOPKPOPKPEEGTSSSEKDEL 109
 RESULT 4
 AAY44959

ID AAY44959 standard; Protein; 115 AA.
 XX
 XX AAY44959;
 AC
 XX 23-MAY-2000 (first entry)
 DT
 XX KDEL receptor inhibitor protein-2.
 DE
 XX
 KW KDEL receptor inhibitor; heat shock protein; immune response;
 KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KW infectious disease; allergy; autoimmune disease.
 XX
 OS Chimeric - Mus sp.
 OS Chimeric - Camelus sp.
 OS Chimeric - Rattus sp.
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..20
 FT /label= Signal peptide
 FT /note= "Derived from mouse Bip"
 FT Domain 26..30
 FT /note= "Altered subsequence from rat COMP"
 FT Domain 30..71
 FT /note= "Rat COMP pentamerisation domain"
 FT Domain 82..105
 FT /note= "Camel IgG linker domain"
 XX
 XX WO200006729-A1.
 XX
 XX 10-FEB-2000.
 XX
 XX 28-JUL-1999; 99WO-US17147.
 XX PF
 XX 29-JUL-1998; 98US-0124671.
 XX
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 PI Rothman JE, Mayhew M, Hoe MH;
 XX
 XX WPI; 2000-195296/17.
 DR N-PSDB; AAZ50493.
 XX
 XX Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX
 XX Disclosure; Fig 2; 87pp; English.
 XX
 CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence is KDEL receptor inhibitor protein comprising regions including
 CC a cleavable signal peptide; the oligomerisation domain from rat
 CC cartilage oligomeric matrix protein (COMP); a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an
 CC alteration of rat COMP which provides increased stability via disulphide
 CC bonds.
 XX
 XX Sequence 115 AA;
 SQ
 Query Match 73.4%; Score 428.5; DB 21; Length 115;
 Best Local Similarity 77.7%; Pred. No. 5.3e-36;
 Matches 87; Conservative 3; Mismatches 5; Indels 17; Gaps 3;

```

QY      8 LLAIAVCSAAKGGSLGDCSDLGPMLELQETNALQDVRLMROQVREITFLKNT 67
      11 LLLLGAV---RAEGSLGSDCC---PQMLELQETNALQDVRLMROQVKEITFLKNT 63
DB      68 VMECDACG-----PQPPKQPQPQPQPQPQPPEPGTSSSEKDEL 109
      64 VMECDACGMQPARTPGTSPPQPPKQPQPQPQPQPPEPGTSSSEKDEL 115

RESULT 5
AA44967
ID      AA44967 standard; Protein, 134 AA.
AC      AA44967;
DT      23-MAY-2000 (first entry)
DE      KDEL receptor inhibitor protein-10.
XX      KDEL receptor inhibitor; heat shock protein; immune response;
KW      oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW      melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW      infectious disease; allergy; autoimmune disease.
XX      Chimeric - Mus sp.
OS      Chimeric - Rattus sp.
OS      Chimeric - Camelus sp.
XX      Key      Location/Qualifiers
      Peptide      1..20
      /label= Signal_peptide
      /note= "Derived from mouse Bip"
      Region      23..32
      /label= Myc_tag_sequence
      /note= "Asn is N-glycosylated"
      Domain      49..94
      /note= "Rat COMP pentamerisation domain"
      Domain      101..124
      /note= "Camel IgG linker domain"
XX      WO200006729-A1.
XX      10-FEB-2000.
XX      28-JUL-1999; 99WO-US17147.
XX      29-JUL-1998; 98US-0124671.
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX      Roelman JE, Mayhew M, Hoe MH;
XX      WPI; 2000-195296/17.
XX      N-PSDB; AA250501.
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
PT      domain useful for promoting secretion of proteins which are normally
PT      retained within the cell -
XX      Disclosure; Fig 10; 87pp; English.
XX      The patent discloses the use of KDEL receptor inhibitor to promote
XX      secretion of proteins that are normally retained within the cell such as
XX      heat shock proteins by inhibiting KDEL receptor-mediated return of
XX      protein complexes to endoplasmic reticulum. This makes the secreted heat
XX      shock proteins more accessible to the immune system and improves immune
XX      response to a target antigen. The inhibitor protein comprises several
XX      subunits where each subunit comprises an oligomerisation domain and has
XX      at its carboxy terminus a region which binds to a KDEL receptor. The
XX      target antigen may be associated with diseases including neoplasia such

```

```

CC      as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
CC      astrocytoma, with defective tumour suppressor genes, oncogenes,
CC      infectious diseases, allergy or autoimmune diseases. The present
CC      sequence encodes KDEL receptor inhibitor comprising regions encoding a
CC      cleavable signal peptide; a myc-tag; an N-glycosylation sequence; the
CC      oligomerisation domain of rat cartilage oligomeric matrix protein
CC      (COMP); a camel IgG linker domain and the carboxy-terminal sequence
CC      KDEL. The subsequence GDCC is an alteration of rat COMP which provides
CC      increased stability via disulphide bonds.
XX      Sequence      134 AA;
SQ      Query Match      72.3%; Score 422; DB 21; Length 134;
      Best local similarity 82.8%; Pred. No. 2.9e-35;
      Matches 82; Conservative 2; Mismatches 1; Indels 14; Gaps 2;
QY      21 GSSLGSDCCSDLGPMLELQETNALQDVRLMROQVREITFLKNTVMECDACG----- 75
      40 GSSLGSDCC---PQMLELQETNALQDVRLMROQVKEITFLKNTVMECDACGMQPAR 95
DB      76 -----PQPPKQPQPQPQPQPQPPEPGTSSSEKDEL 109
      96 TPGRSPQPPKQPQPQPQPQPQPPEPGTSSSEKDEL 134
QY      RESULT 6
      AA44958
      ID      AA44958 standard; Protein, 115 AA.
      AC      AA44958;
      DT      23-MAY-2000 (first entry)
      DE      KDEL receptor inhibitor protein-1.
      XX      KDEL receptor inhibitor; heat shock protein; immune response;
      KW      oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
      KW      melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
      KW      infectious disease; allergy; autoimmune disease.
      XX      Chimeric - Mus sp.
      OS      Chimeric - Camelus sp.
      OS      Chimeric - Rattus sp.
      XX      Key      Location/Qualifiers
      Peptide      1..20
      /label= Signal_peptide
      /note= "Derived from mouse Bip"
      Domain      26..30
      /note= "Subsequence from rat COMP"
      Domain      30..71
      /note= "Rat COMP pentamerisation domain"
      Domain      82..105
      /note= "Camel IgG linker domain"
XX      WO200006729-A1.
XX      10-FEB-2000.
XX      28-JUL-1999; 99WO-US17147.
XX      29-JUL-1998; 98US-0124671.
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX      Roelman JE, Mayhew M, Hoe MH;
XX      WPI; 2000-195296/17.
XX      N-PSDB; AA250492.
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
PT      domain useful for promoting secretion of proteins which are normally
PT      retained within the cell -

```

XX PS Disclosure; Fig 1; 87pp; English.

CC The patent discloses the use of KDEL receptor inhibitor to promote

CC secretion of proteins that are normally retained within the cell such as

CC heat shock proteins by inhibiting KDEL receptor-mediated return of

CC protein complexes to endoplasmic reticulum. This makes the secreted heat

CC shock proteins more accessible to the immune system and improves immune

CC response to a target antigen. The inhibitor protein comprises several

CC subunits where each subunit comprises an oligomerisation domain and has

CC at its carboxy terminus a region which binds to a KDEL receptor. The

CC target antigen may be associated with diseases including neoplasia such

CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and

CC astrocytoma, with defective tumour suppressor genes, oncogenes,

CC infectious diseases, allergy or autoimmune diseases. The present

CC sequence is KDEL receptor inhibitor protein comprising regions including

CC a cleavable signal peptide; the oligomerisation domain from rat

CC cartilage oligomeric matrix protein; a camel IgG linker domain and the

CC carboxy-terminal sequence KDEL.

XX SQ Sequence 115 AA;

Query Match 71.0%; Score 414.5; DB 21; Length 115;

Best Local Similarity 76.8%; Pred. No. 1.4e-34;

Matches 86; Conservative 3; Mismatches 6; Indels 17; Gaps 3;

Oy 8 LIALAAVCSAKKSSSLGDCSPDLGPGMLRELOETNALQDVDMRLRQVREITFLKNT 67

Db 11 LLLLGAV---RAEGSSLOG---DLAPQMLELOETNALQDVVELLRQVKEITFLKNT 63

Oy 68 VMECDACG-----POPQPKPOPOPQPKPOPKPEPEGSSSEKDEL 109

Db 64 VMECDACGMOPARTPTGTSPPQPKPOPKPOPKPEPEGSSSEKDEL 115

RESULT 7

AAV44962

ID AAV44962 standard; Protein; 109 AA.

XX AAV44962;

XX 23-MAY-2000 (first entry)

XX KDEL receptor inhibitor protein-5.

XX KDEL receptor inhibitor; heat shock protein; immune response;

XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;

XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

XX infectious disease; allergy; autoimmune disease.

XX Chimeric - Mus sp.

OS Chimeric - Xenopus sp.

OS Chimeric - Camelus sp.

OS Chimeric - Rattus sp.

XX Key location/Qualifiers

FT Peptide 1..20

FT /label= Signal peptide

FT /note= "Derived from mouse B1p"

FT Domain 26..30

FT /note= "Altered subsequence from rat COMP"

FT Domain 30..75

FT /note= "Xenopus thrombospondin 4 trimerisation domain"

FT Domain 76..99

FT /note= "Camel IgG linker domain"

XX MO200006729-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17147.

XX 29-JUL-1998; 98US-0124671.

XX PA (SLOC) SLOAN KETTERING INST CANCER RES.

XX P1 Rothman JE, Mayhew M, Hoe MH;

XX WP1; 2000-195296/17.

DR N-PSDB; AAZ50496.

XX Inhibitors of the KDEL receptor which comprises an oligomerization

PT domain useful for promoting secretion of proteins which are normally

PT retained within the cell

XX Disclosure; Fig 5; 87pp; English.

CC The patent discloses the use of KDEL receptor inhibitor to promote

CC secretion of proteins that are normally retained within the cell such as

CC heat shock proteins by inhibiting KDEL receptor-mediated return of

CC protein complexes to endoplasmic reticulum. This makes the secreted heat

CC shock proteins more accessible to the immune system and improves immune

CC response to a target antigen. The inhibitor protein comprises several

CC subunits where each subunit comprises an oligomerisation domain and has

CC at its carboxy terminus a region which binds to a KDEL receptor. The

CC target antigen may be associated with diseases including neoplasia such

CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and

CC astrocytoma, with defective tumour suppressor genes, oncogenes,

CC infectious diseases, allergy or autoimmune diseases. The present

CC sequence is KDEL receptor inhibitor protein comprising regions encoding a

CC cleavable signal peptide; the oligomerisation domain of Xenopus

CC thrombospondin 4 (TSP4) trimerisation domain including an additional

CC subsequence; a camel IgG linker domain and the carboxy-terminal sequence

CC KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric

CC matrix protein which provides increased stability via disulphide bonds.

XX SQ Sequence 109 AA;

Query Match 63.6%; Score 371.5; DB 21; Length 109;

Best Local Similarity 66.7%; Pred. No. 3.2e-30;

Matches 68; Conservative 11; Mismatches 18; Indels 3; Gaps 1;

Oy 8 LIALAAVCSAKKSSSLGDCSPDLGPGMLRELOETNALQDVDMRLRQVREITFLKNT 67

Db 11 LLLLGAV---RAEGSSSLGDCSPDLGPGMLRELOETNALQDVDMRLRQVREITFLKNT 67

Oy 68 VMECDACGPOPOPQPKPOPOPQPKPOPKPEPEGSSSEKDEL 109

Db 68 IAEQACGPOPOPQPKPOPOPQPKPOPKPEPEGSSSEKDEL 109

RESULT 8

AAV44961

ID AAV44961 standard; Protein; 109 AA.

XX AAV44961;

XX 23-MAY-2000 (first entry)

XX KDEL receptor inhibitor protein-4.

XX KDEL receptor inhibitor; heat shock protein; immune response;

XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;

XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

XX infectious disease; allergy; autoimmune disease.

XX Chimeric - Mus sp.

OS Chimeric - Xenopus sp.

OS Chimeric - Camelus sp.

OS Chimeric - Rattus sp.

XX Key location/Qualifiers

FT Peptide 1..20

FT /label= Signal peptide

FT /note= "Derived from mouse B1p"

FT Domain 26..30

FT /note= "Altered subsequence from rat COMP"

FT	Domain		30..75	/note= "Mouse thrombospondin 3 trimerisation domain"
FT			76..99	
FT	Domain		/note= "Camel igg linker domain"	
XX				
XX				
XX				
PA	(SLOK) SLOAN KETTERING INST CANCER RES.			
PR	29-JUL-1998;	98US-0124671.		
PF	28-JUL-1999;	99WO-US17147.		
PB	10-FEB-2000.			
PD				
PN	WO200006729-A1.			
PP				
PT	Inhibitors of the KDEL receptor which comprises an oligomerization domain useful for promoting secretion of proteins which are normally retained within the cell -			
PS	Disclosure; Fig 4; 87pp; English.			
XX				
CC	The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several substituents where each substituent comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergy or autoimmune diseases. The present sequence is KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; the oligomerisation domain of mouse thrombospondin 3 (TSP3) trimerisation domain including an additional subsequence; a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric matrix protein which provides increased stability via disulphide bonds.			
SQ	Sequence	109 AA;		
Query Match		59.2%; Score 345.5; DB 21; Length 109;		
Best Local Similarity		62.7%; Pred. No. 1.4e-21; Indels 3; Gaps		
Matches	64; Conservative 14; Mismatches 21;			
DY	8 LTLALAAVCAAKGSSLSIGDCCSDLPOMLRLEQETNAALQDVRLRQGVREITFLKNT	67		
DB	11 LLLLGAV---RAEGSSISGDGCCEGTQFVALVTGLTFNQLIVELRDDIRIDQVEMSLIRNT	67		
OY	68 VMEEDACGPORPKPQPQRPOPKPEPEEETGSSEKDEL	109		
DB	68 IMEQVCGPQPQPKPQPQPQPKPEPEEETGSSEKDEL	109		
RESULT 9				
AAY44960	ID AAY44960 standard; Protein: 105 AA.			
XX				
AC	AAY44960;			
XX				
DT	23-MAY-2000 (first entry)			
XX				
DE	KDEL receptor inhibitor protein-3.			
KM	KDEL receptor inhibitor; heat shock protein; immune response;			
KW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;			
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;			

KW		infectious disease; allergy; autoimmune disease.
XX		
OS	Chimeric - Mus sp.	
OS	Chimeric - Camelus sp.	
OS	Chimeric - Rattus sp.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..20
FT	/label= Signal peptide	
FT	/note= "Derived from mouse BIP"	
FT	Domain	26..30
FT	/note= "Altered subsequence from rat COMP"	
FT	Domain	30..71
FT	/note= "Mouse thrombospondin 3 trimerisation domain"	
FT	Domain	72..95
FT	/note= "Camel IgG linker domain"	
PV		
NV	WO200006729-A1.	
PD	10-FEB-2000.	
PP	28-JUL-1999;	99MO-USI7147.
PR	29-JUL-1998;	98US-O124671.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.	
PI	Rothman JE, Mayhew M, Hoe MH;	
XX		
DR	WPt ; 2000-195296/17.	
NR	N-PSTDB; AAZ50494.	
PT	Inhibitors of the KDEL receptor which comprises an oligomerization domain useful for promoting secretion of proteins which are normally retained within the cell _	
PS		
XX	Disclosure; Fig 3; 87bp; English.	
CC	The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several subunits where each subunit comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergic or autoimmune diseases. The present sequence is KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; the oligomerisation domain of mouse thrombospondin 3 (TPP3) trimerisation domain; a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric matrix protein which provides increased stability via disulphide bonds.	
SQ	Sequence 105 AA;	
OY	Query Match 57.8%; Score 337.5; DB 21; Length 105; Best Local Similarity 63.7%; Pred. No. 9e-47;	
Dz	Matches 65; Conservative 12; Mismatches 18; Indels 7; Gaps 2	
OY	8 LLAIAVCSAKKSSIGGCCSDLGFMELRLEDTNNAADVDMDLRQQREITFLKYT 67	: :
Dz	11 LLILGAV---RAEBSISGGDCCKAL---VTQLTLFNOILVELRDDRIDDKVEMSLIINT 63	: :
OY	VMECDACGPQPQPOPQOPQRPQRPEEGTGSSSEKDEL 109	
Dz	64 IMECVGCGPGQPKFPQPOPPQPKPFEPFGTGSSEKDEL 105	::::

AAV44964
ID AAV44964 standard; Protein; 90 AA.
XX
AC AAV44964;
XX
DT 23-MAY-2000 (first entry)
XX
DE KDEL receptor inhibitor protein-7.
XX
KW KDEL receptor inhibitor; heat shock protein; immune response;
KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW infectious disease; allergy; autoimmune disease.
XX
OS Chimeric - Adenovirus B3.
OS Chimeric - Homo sapiens.
OS Chimeric - Camelus sp.
XX
Key Location/Qualifiers
FH Peptide 1..20
FT /note= "Derived from adenovirus B3"
FT Domain 30..56
FT /note= "Human phospholamban pentamerisation domain"
FT Domain 57..80
FT /note= "Camel IgG linker domain"
XX
PN WO200006729-A1.
XX
PD 10-FEB-2000.
XX
PF 28-JUL-1999; 99WO-US17147.
XX
PR 29-JUL-1998; 98US-0124671.
XX
PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX
PI Rothman JE, Mayhew M, Hoe MH;
XX
DR WPI: 2000-195296/17.
DR N-PSDB; AAZ50498.
XX
PT Inhibitors of the KDEL receptor which comprises an oligomerization
PT domain useful for promoting secretion of proteins which are normally
PT retained within the cell -
XX
PS Disclosure; Fig 7; 87pp; English.
XX
CC The patent discloses the use of KDEL receptor inhibitor to promote
CC secretion of proteins that are normally retained within the cell such as
CC heat shock proteins by inhibiting KDEL receptor-mediated return of
CC protein complexes to endoplasmic reticulum. This makes the secreted heat
CC shock proteins more accessible to the immune system and improves immune
CC response to a target antigen. The inhibitor protein comprises several
CC subunits where each subunit comprises an oligomerisation domain and has
CC at its carboxy terminus a region which binds to a KDEL receptor. The
CC target antigen may be associated with diseases including neoplasia such
CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
CC astrocytoma, with defective tumour suppressor genes, oncogenes,
CC infectious diseases, allergy or autoimmune diseases. The present
CC sequence is KDEL receptor inhibitor comprising regions encoding a
CC cleavable signal peptide; the oligomerisation domain of human
CC phospholamban (PLB) pentamerisation domain; a camel IgG linker domain
CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an
CC alteration of rat cartilage oligomeric matrix protein which provides
CC increased stability via disulphide bonds.
XX
SQ Sequence 90 AA;
XX
Query Match 53.0%; Score 309.5; DB 21; Length 90;
Best Local Similarity 59.6%; Pred. No. 5,3e-24;
Matches 68; Conservative 1; Mismatches 16; Indels 29; Gaps 2;
QY 1 MRYMILGLALAAVCSAAKKGSSLGDCSDLPQMLRELOETNAALQDVWDMLRQVRE 60

DB 1 MRYMILGLALAAVCSAAKKGSSLGDCSDLPQMLRELOETNAALQDVWDMLRQVRE 109
QY 61 ITFLKNTVMECDAC-----GQPPQPKPQPPQPPQPKPKEPCTGSSSEDEL 109
DB 37 INFCLILICLLILICILVWLLPQPPKPPQPPQPPKPPKPEPCTGSSSEDEL 90
RESULT 11
AAB00040
ID AAB00040 standard; Protein; 242 AA.
XX
AC AAB00040;
XX
DT 08-NOV-2000 (first entry)
XX
DE Human COMP/TSP-1 chimeric protein.
XX
DE TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KW thrombospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
KW glaucoma.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2000044908-A2.
XX
PD 03-AUG-2000.
XX
PF 01-FEB-2000; 2000WO-US02482.
XX
PR 01-FEB-1999; 99US-0118053.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lawler JW;
XX
DR WPI: 2000-514823/46.
DR N-PSDB; AAA47734.
XX
PT Nucleic acids encoding chimeric proteins such as cartilage oligomeric
PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for
PT inhibiting angiogenesis and treating diseases such as cancer
XX
PS Claim 33; Fig 4a-b; 40pp; English.
XX
CC New nucleic acids are described which encode a protein comprising
CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
CC but not the TGF (transforming growth factor)-beta activation region
CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
CC the second and third type-1 repeats and the COMP (cartilage
CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
CC caused inhibition of the growth of tumours in mice models.
CC Thus the nucleic acids and proteins may be useful for treating
CC angiogenesis related diseases such as cancer (by reducing the rate of
CC growth and size of tumours), arthritis, psoriasis, diabetic
CC retinopathy, corneal graft rejection, and glaucoma. They may also be
CC used for treating human immunodeficiency virus (HIV) infection.
CC Anti-angiogenic therapy has little toxicity, does not require the
CC therapeutic agent to enter tumour cells or cross the blood-brain
CC barrier, controls tumour growth independently of growth of
CC tumour cell heterogeneity, and does not induce drug resistance.
XX
SQ Sequence 242 AA;
XX
Query Match 44.9%; Score 262.5; DB 21; Length 242;
Best Local Similarity 69.5%; Pred. No. 1e-18; Indels 11; Gaps 2;
Matches 57; Conservative 5; Mismatches 9;
QY 8 LLAALAAVCSAAKKGSSLGDCSDLPQMLRELOETNAALQDVWDMLRQVREITFLKNT 67

```

Db      10 LITLALGASGGGSPLG---SDLGPMRELQETNALQDVRDMLRQVREITFLKNT 65
QY      68 VMECDACGPQ-----PQPKP 82
        |||||
        |
Db      66 VMECDACGMQGSVRTGLPSVRP 87

RESULT 12
AAB00041 ID AAB00041 standard; Protein; 300 AA.
XX
AC AAB00041;
XX
DT 08-NOV-2000 (first entry)
XX
DE Human COMP/TSP-2 chimeric protein.
XX
KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KM chromospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
KM glaucoma.
XX
OS Homo sapiens.
XX
XX Synthetic.
XX
PN WO200044908-A2.
XX
PD 03-AUG-2000.
XX
PF 01-FEB-2000; 2000WO-US02482.
XX
PR 01-FEB-1999; 99US-0118053.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lawler JW;
XX
DR WPI; 2000-514823/46.
XX
N-PSDB; AAA47735.
XX
PT Nucleic acids encoding chimeric proteins such as cartilage oligomeric
PT matrix protein (COMP)/chromospondins (TSP)-1 and 2, useful for
PT inhibiting angiogenesis and treating diseases such as cancer
XX
XX Claim 50; Fig 5a-b; 40pp; English.
XX
PS New nucleic acids are described which encode a protein comprising
CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
CC but not the TGF (transforming growth factor)-beta activation region
CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
CC the second and third type-1 repeats and the COMP (cartilage
CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
CC caused inhibition of the growth of tumours in mice models.
CC Thus the nucleic acids and proteins may be useful for treating
CC angiogenesis related diseases such as cancer (by reducing the rate of
CC growth and size of tumours), arthritis, psoriasis, diabetic
CC retinopathy, corneal graft rejection, and glaucoma. They may also be
CC used for treating human immunodeficiency virus (HIV) infection.
CC Anti-angiogenic therapy has little toxicity, does not require the
CC therapeutic agent to enter tumour cells or cross the blood-brain
CC barrier, controls tumour growth independently of growth of
CC tumour cell heterogeneity, and does not induce drug resistance.
XX
SQ Sequence 300 AA:

Query Match 44.9%; Score 262.5; DB 21; Length 300;
Best Local Similarity 69.5%; Pred. No. 1,4e-18;
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

      8 LITLALGASAAKKSSSLGGCCSDLGPMRELQETNALQDVRDMLRQVREITFLKNT 67
      |||||
      |
      10 LITLALGASGGGSPLG---SDLGPMRELQETNALQDVRDMLRQVREITFLKNT 65

```

```

QY      68 VMECDACGPQ-----PQPKP 82
        |||||
        |
Db      66 VMECDACGMQGSVRTGLPSVRP 87

RESULT 13
AAB00044 ID AAB00044 standard; Protein; 757 AA.
XX
AC AAB00044;
XX
DT 08-NOV-2000 (first entry)
XX
DE Human cartilage oligomeric matrix protein (COMP).
XX
KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
KM chromospondin; angiogenesis; tumour; treatment; cancer;
KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
KM glaucoma.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH Region 89..128
XX FT /label= Type 2 repeat region
XX FT 129..181
XX FT /label= Type 2 repeat region
XX FT 182..226
XX FT /label= Type 2 repeat region
XX FT 227..268
XX FT /label= Type 2 repeat region
XX
PN WO200044908-A2.
XX
PD 03-AUG-2000.
XX
PF 01-FEB-2000; 2000WO-US02482.
XX
PR 01-FEB-1999; 99US-0118053.
XX
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI Lawler JW;
XX
DR WPI; 2000-514823/46.
XX
XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric
XX matrix protein (COMP)/chromospondins (TSP)-1 and 2, useful for
XX inhibiting angiogenesis and treating diseases such as cancer
XX
XX Disclosure; Fig 3; 40pp; English.
XX
PS New nucleic acids are described which encode a protein comprising
CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
CC but not the TGF (transforming growth factor)-beta activation region
CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
CC the second and third type-1 repeats and the COMP (cartilage
CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
CC caused inhibition of the growth of tumours in mice models.
CC Thus the nucleic acids and proteins may be useful for treating
CC angiogenesis related diseases such as cancer (by reducing the rate of
CC growth and size of tumours), arthritis, psoriasis, diabetic
CC retinopathy, corneal graft rejection, and glaucoma. They may also be
CC used for treating human immunodeficiency virus (HIV) infection.
CC Anti-angiogenic therapy has little toxicity, does not require the
CC therapeutic agent to enter tumour cells or cross the blood-brain
CC barrier, controls tumour growth independently of growth of
CC tumour cell heterogeneity, and does not induce drug resistance.
XX
SQ Sequence 757 AA:

```


Query Match 44.9%; Score 262.5; DB 21; Length 757;
 Best Local Similarity 69.5%; Pred. No. 4e-18;
 Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LIALAAVCSAAKKSSSLGDCCSDLGPQMLRELOETNAALQDVWDMLRQVREITFLKNT 67
 DB 10 LTLTAAAGASQGSGSPLG---SDLGPQMLRELOETNAALQDVWDMLRQVREITFLKNT 65
 QY 68 VMECDACGPO-----PQPKP 82
 DB 66 VMECDACGMOQSVRTGLPSVRP 87

RESULT 14
 ID ABR05594 standard; Protein; 757 AA.
 AC ABR05594;
 DT 14-NOV-2002 (first entry)
 XX Breast cancer-associated protein 59.
 DE Breast cancer-associated protein 59.
 XX Breast cancer; breast cancer-associated gene sequence;
 KM drug development; pharmacogenetics; biosensor development.
 XX Unidentified.
 OS
 XX NO200259377-A2.
 PN
 XX 01-AUG-2002.
 PD
 XX 24-JAN-2002; 2002WO-US02242.
 PF
 XX 24-JAN-2001; 2001US-263965P.
 PR 02-FEB-2001; 2001US-265928P.
 PR 09-APR-2001; 2001US-282698P.
 PR 09-APR-2001; 2001US-082947Z.
 PR 04-MAY-2001; 2001US-288590P.
 PR 29-MAY-2001; 2001US-294443P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 PI Mack DH, Gish KC, Afar D;
 DR WPI: 2002-583738/62.
 XX N-PSDB; ABR07751.
 XX
 PT Detecting a breast cancer-associated transcript in a patient's cell,
 PT useful for diagnosing breast cancer, comprises contacting a biological
 PT sample with a polynucleotide that selectively hybridizes with breast
 PT cancer nucleic acids -
 PS
 XX Disclosure; Page 401; 414pp; English.

The invention comprises a method of detecting a breast cancer-associated transcript in a cell from a patient. The method of the invention involves contacting a biological sample from the patient with a nucleotide that hybridizes to one of the 69 breast cancer-associated gene sequences shown in the specification. The method of the invention is useful in the diagnosis or prognosis of breast cancer, and for detecting genes that are up or down-regulated in breast cancer cells. Genes identified by the method of the invention can be used in diagnostic purposes and also as targets for screening for therapeutic compounds that modulate breast cancer (e.g. hormones or antibodies). Identification of genes that are over or under expressed in breast cancer can additionally provide high-resolution, high-sensitivity datasets which can be used in the areas of diagnostics, therapeutics, drug development, pharmacogenetics, protein structure and biosensor development. Amino acid sequences ABR05536 - ABR05604 represent the proteins encoded by the 69 breast cancer-associated genes of the invention.

Sequence 757 AA;
 SQ

Query Match 44.9%; Score 262.5; DB 23; Length 757;
 Best Local Similarity 69.5%; Pred. No. 4e-18;
 Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LIALAAVCSAAKKSSSLGDCCSDLGPQMLRELOETNAALQDVWDMLRQVREITFLKNT 67
 DB 10 LTLTAAAGASQGSGSPLG---SDLGPQMLRELOETNAALQDVWDMLRQVREITFLKNT 65
 QY 68 VMECDACGPO-----PQPKP 82
 DB 66 VMECDACGMOQSVRTGLPSVRP 87

RESULT 15
 ID ABR47420 standard; Protein; 757 AA.
 AC ABR47420;
 DT 12-JUN-2003 (first entry)
 XX Breast cancer associated protein sequence SEQ ID NO:72.
 DE Breast cancer associated protein sequence SEQ ID NO:72.
 XX Human; breast cancer; cytostatic; gene therapy.
 KM
 XX Homo sapiens.
 OS
 XX MO2003004989-A2.
 PN
 XX 16-JAN-2003.
 PD
 XX 21-JUN-2002; 2002WO-US19669.
 PF
 XX 21-JUN-2001; 2001US-299887P.
 PR 27-JUN-2001; 2001US-301572P.
 PR 18-JUL-2001; 2001US-306501P.
 PR 25-SEP-2001; 2001US-325002P.
 PR 05-MAR-2002; 2002US-362585P.
 PR 14-MAY-2002; 2002US-380391P.
 XX
 PA (MILL-) MILLENIUM PHARM INC.
 PI Lillie J, Gamavaranu M, Glatc K, Hoerath S, Kamatkar S, Mertens M;
 PI Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Bast RC;
 PI Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
 XX
 DR WPI: 2003-210381/20.
 DR N-PSDB; ACC50112.
 XX
 PT Breast cancer diagnosis or treatment by comparing the level of
 PT expression of a marker in a patient sample with that in the control
 PT non-breast cancer sample -
 PS
 XX Claim 1; SEQ ID 72; 128pp; English.

The present invention describes a method for assessing whether a patient is afflicted with breast cancer. The method comprises comparing the level of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and ABR47386 to ABR47632) in a patient sample and the normal level of expression of the marker in a control non-breast cancer sample, where a significant increase in the level of expression of the marker in the patient sample and the normal level is an indication that the patient is afflicted with breast cancer. The breast cancer associated sequences from the present invention have cytostatic activities and can be used in gene therapy. The method is useful for diagnosing and treating breast cancer.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 757 AA;
 SQ

Query Match 44.9%; Score 262.5; DB 24; Length 757;
Best Local Similarity 69.5%; Pred. No. 4e-18;
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;
QY 8 LLALAAVCSAAKGGSSLGDDCCSDLGPMRLRELOETNAALQDVRDMLRQGVREITFLKNT 67
|||: : : |||
Db 10 LITLALAGASGGGSPFG---SDLGPMRLRELOETNAALQDVRDMLRQGVREITFLKNT 65
|||: : : |||
QY 68 VMECDACGPGQ-----PQPKP 82
|||: : : |||
Db 66 VMECDACGMQGSVRTGLPSVRP 87
|||: : : |||

Search completed: August 26, 2003, 16:04:22
Job time : 118.717 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:05:48 ; Search time 38.5841 Seconds
(without alignments)
119.528 Million cell updates/sec

Title: US-09-696-872-23

Perfect score: 584
Sequence: 1 MRWILGLLALAAVCSAAKK.....PKQPKPEECTGSSSEKDEL 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	584	100.0	109	US-09-124-671-23	Sequence 23, Appl
2	464	79.5	109	US-09-124-671-29	Sequence 29, Appl
3	434	74.3	109	US-09-124-671-27	Sequence 27, Appl
4	428.5	73.4	115	US-09-124-671-15	Sequence 15, Appl
5	422	72.3	134	US-09-124-671-34	Sequence 34, Appl
6	414.5	71.0	115	US-09-124-671-13	Sequence 13, Appl
7	371.5	63.6	109	US-09-124-671-21	Sequence 21, Appl
8	345.5	59.2	109	US-09-124-671-19	Sequence 19, Appl
9	337.5	57.8	105	US-09-124-671-17	Sequence 17, Appl
10	309.5	53.0	90	US-09-124-671-25	Sequence 25, Appl
11	241	41.3	46	US-09-124-671-2	Sequence 2, Appl
12	234.5	40.2	66	US-09-091-814-46	Sequence 46, Appl
13	228	39.0	64	US-09-091-814-45	Sequence 45, Appl
14	211	36.1	46	US-09-124-671-1	Sequence 1, Appl
15	146	25.0	35	US-08-471-780C-38	Sequence 38, Appl
16	146	25.0	35	US-08-467-282B-38	Sequence 38, Appl
17	146	25.0	35	US-08-471-282A-38	Sequence 38, Appl
18	146	25.0	35	US-08-466-710C-38	Sequence 38, Appl
19	146	25.0	35	US-08-468-739C-38	Sequence 38, Appl
20	146	25.0	35	US-08-471-780C-44	Sequence 44, Appl
21	146	25.0	54	US-08-467-282B-44	Sequence 44, Appl
22	146	25.0	54	US-08-471-282A-44	Sequence 44, Appl
23	146	25.0	54	US-08-466-710C-44	Sequence 44, Appl
24	146	25.0	54	US-08-468-739C-44	Sequence 44, Appl
25	146	25.0	60	US-08-471-780C-87	Sequence 87, Appl
26	146	25.0	60	US-08-467-282B-87	Sequence 87, Appl
27	146	25.0	60	US-08-471-282A-87	Sequence 87, Appl

28	146	25.0	60	US-08-466-710C-87	Sequence 87, Appl
29	146	25.0	60	US-08-468-739C-87	Sequence 87, Appl
30	145.5	24.9	961	PCT-US93-11725-4	Sequence 4, Appl
31	144	24.7	24	US-09-379-297-5	Sequence 5, Appl
32	138.5	23.7	889	PCT-US93-11725-2	Sequence 2, Appl
33	121	20.7	46	US-09-124-671-5	Sequence 5, Appl
34	119	20.4	667	US-08-718-661-2	Sequence 2, Appl
35	117	20.0	46	US-09-124-671-6	Sequence 6, Appl
36	116	19.9	261	US-09-602-565-34	Sequence 34, Appl
37	112.5	19.3	907	US-08-989-299-12	Sequence 12, Appl
38	111	19.0	221	US-09-069-023-22	Sequence 22, Appl
39	109	18.7	558	US-09-252-991A-17202	Sequence 17202, A
40	108	18.5	304	US-09-252-991A-23116	Sequence 23116, A
41	107	18.3	613	US-09-345-473B-39	Sequence 39, Appl
42	102	17.5	8991	US-08-714-741-32	Sequence 32, Appl
43	101	17.3	360	US-09-252-991A-19088	Sequence 19088, A
44	99	17.0	694	US-09-328-352-7276	Sequence 7276, Ap
45	98	16.8	619	US-08-465-746-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-124-671-23
; Sequence 23, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human COMP-KDEL
US-09-124-671-23

Query Match      100.0%; Score 584; DB 3; Length 109;
Best Local Similarity 100.0%; Pred. No. 2.2e-57;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 MRWILGLLALAAVCSAAKKSSISGDCSDLPQMRLEIQTNALQDVDMVLRQVRE 60
Db      1 MRWILGLLALAAVCSAAKKSSISGDCSDLPQMRLEIQTNALQDVDMVLRQVRE 60

Cy      61 ITFLKNTVMCDACGPPQPKPPQPPQPKPPQPKPEECTGSSSEKDEL 109
Db      61 ITFLKNTVMCDACGPPQPKPPQPPQPKPPQPKPEECTGSSSEKDEL 109

RESULT 2
US-09-124-671-29
; Sequence 29, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
```

```

; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human TSP4-KDEL
US-09-124-671-29

Query Match
Best Local Similarity 79.5%; Score 464; DB 3; Length 109;
Matches 85; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 1 MRYMILGLIALAAVCSAAKGGSLGGDCSDLGPMRLRELOETNAALQDVRLMROQVRE 60
DB 1 MRYMILGLIALAAVCSAAKGGSLGGDCSDLGPMRLRELOETNAALQDVRLMROQVRE 60

QY 61 ITFLKNTWECDCGPGPOPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 109
DB 61 TSFLNRTIAECQACGPGPOPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 109

RESULT 3
US-09-124-671-27
; Sequence 27, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human TSP3-KDEL
US-09-124-671-27

Query Match
Best Local Similarity 74.3%; Score 434; DB 3; Length 109;
Matches 79; Conservative 13; Mismatches 17; Indels 0; Gaps 0;

QY 1 MRYMILGLIALAAVCSAAKGGSLGGDCSDLGPMRLRELOETNAALQDVRLMROQVRE 60
DB 1 MRYMILGLIALAAVCSAAKGGSLGGDCSDLGPMRLRELOETNAALQDVRLMROQVRE 60

QY 61 ITFLKNTWECDCGPGPOPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 109
DB 61 MSLINTIMECQVCGPGPOPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 109

RESULT 4
US-09-124-671-15
; Sequence 15, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```

; FEATURE:
; OTHER INFORMATION: chimeric rat COMP-KDEL
US-09-124-671-15

Query Match
Best Local Similarity 73.4%; Score 428.5; DB 3; Length 115;
Matches 87; Conservative 3; Mismatches 5; Indels 17; Gaps 3;

QY 8 LIALAAVCSAAKGGSLGGDCSDLGPMRLRELOETNAALQDVRLMROQVREITFLKNT 67
DB 11 LILGLAV---RAESSLSGGDC---PQMLRELOETNAALQDVRELRQGVKEITFLKNT 63

QY 68 VMECDACG-----PQPKPOPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 109
DB 64 VMECDACGMQPARTPGTSPPQPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 115

RESULT 5
US-09-124-671-34
; Sequence 34, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: KDEL/myc
US-09-124-671-34

Query Match
Best Local Similarity 72.3%; Score 422; DB 3; Length 134;
Matches 82; Conservative 2; Mismatches 1; Indels 14; Gaps 2;

QY 21 GSSLSGGDCSDLGPMRLRELOETNAALQDVRLMROQVREITFLKNTVMECDACG----- 75
DB 40 GSSLSGGDC---PQMLRELOETNAALQDVRELRQGVKEITFLKNTVMECDACGMQPAR 95

QY 76 -----PQPKPOPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 109
DB 96 TPGTSPQPKPOPKPOPKPOPKPOPKPEPEGTGSSSEKDEL 134

RESULT 6
US-09-124-671-13
; Sequence 13, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric rat comp
US-09-124-671-13

```

Query Match	53.0%	Score	309.5	DB	3	Length	90
Best Local Similarity	59.6%	Pred	No. 4.1e-27				
Matches	68	Conservative	1	Mismatches	16	Indels	29
						Gaps	2
Y	1	MRYYIGLILAAVCAAKKSSLGSGDCCSDLGPQMRLEIDETNAALQDVDMRLQVRE	60				


```

1      CITY: Washington
2      STATE: D C
3      COUNTRY: USA
4      ZIP: 20005-3315
5      COMPUTER READABLE FORM:
6      MEDIUM TYPE: Floppy disk
7      COMPUTER: IBM PC compatible
8      OPERATING SYSTEM: PC-DOS/MS-DOS
9      SOFTWARE: Patentin Release #1.0, Version #1.25
10     CURRENT APPLICATION DATA:
11     APPLICATION NUMBER: US/08/471,780C
12     FILING DATE: 06-JUN-1995
13     CLASSIFICATION: 530
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: US 08/106,944
16     FILING DATE: 17-AUG-1993
17     APPLICATION NUMBER: FR 92402326.0
18     FILING DATE: 21-AUG-1992
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER: FR 93401310.3
21     FILING DATE: 21-MAY-1993
22     ATTORNEY/AGENT INFORMATION:
23     NAME: Potter, Jane E.R.
24     REGISTRATION NUMBER: 33,332
25     REFERENCE/DOCKET NUMBER: 04958.0008-00000
26     TELECOMMUNICATION INFORMATION:
27     TELEPHONE: 202-408-4000
28     TELEFAX: 202-408-4400
29     INFORMATION FOR SEQ ID NO: 38:
30     SEQUENCE CHARACTERISTICS:
31     LENGTH: 35 amino acids
32     TYPE: amino acid
33     STRANDEDNESS: single
34     TOPOLOGY: linear
35     FEATURE:
36     NAME/KEY: Region
37     LOCATION: 1..35
38     OTHER INFORMATION: /label= hinge
39
40     US-08-471-780C-38

```

		Query Match	25.0%	Score 146;	DB 1;	Length 35;
		Best Local Similarity	96.2%;	Pred.	No. 1.6e-09;	
		Matches	25;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
QY	76	PQQPQPKPPQPQQPQPKPPKPEECT	101			
Db	5	PQQPQKPQPPQPQQPQPKPPKPEECT	30			

Search completed: August 26, 2003, 16:19:15
Job time : 39.5841 secs

This Page Blank (unsp'd)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:02:18 ; Search time 38.5841 Seconds
(without alignments)
372.792 Million cell updates/sec

Title: US-09-696-872-23

Perfect score: 584
Sequence: 1 MRWIIIGLALAAVCSAAK.....PKPQKPEECTGSSEKDEL 109

Scoring table: BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 497079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262.5	44.9	242	10	US-09-919-603-5
2	262.5	44.9	300	10	US-09-919-603-7
3	262.5	44.9	757	10	US-09-919-603-3
4	262.5	44.9	757	12	US-10-301-823-41
5	262.5	44.9	757	15	US-10-177-293-72
6	145.5	24.9	961	15	US-10-017-721-4
7	140	24.0	178	15	US-10-154-971-22
8	139	23.8	303	15	US-10-154-971-24
9	126	21.6	578	15	US-10-156-761-12544
10	122	20.9	400	12	US-10-301-823-155
11	116	19.9	261	15	US-10-280-953-17
12	115	19.7	285	10	US-09-881-752A-228
13	111	19.0	807	14	US-10-138-221-7
14	108	18.5	316	10	US-09-764-864-967
15	107.5	18.4	266	10	US-09-738-626-6457

16	107	18.3	613	10	US-09-862-027-39	Sequence 39, Appl
17	105	18.0	265	14	US-10-073-256-78	Sequence 78, Appl
18	102	17.5	417	12	US-09-949-029-2	Sequence 2, Appl
19	102	17.5	1333	9	US-09-815-242-10936	Sequence 10936, A
20	99	17.0	54	10	US-09-855-754-22	Sequence 22, Appl
21	99	17.0	827	15	US-10-171-384-3	Sequence 3, Appl
22	98	17.0	1493	9	US-09-858-754-4	Sequence 4, Appl
23	98	16.8	619	11	US-09-882-774-1	Sequence 1, Appl
24	97	16.6	250	10	US-09-252-150-20	Sequence 20, Appl
25	97	16.6	583	15	US-10-156-761-12356	Sequence 12356, A
26	95.5	16.4	605	10	US-09-331-631A-25	Sequence 25, Appl
27	94.5	16.2	183	10	US-09-252-150-2	Sequence 2, Appl
28	94	16.1	59	10	US-09-855-754-16	Sequence 16, Appl
29	94	16.1	762	11	US-09-917-378-1	Sequence 1, Appl
30	94	16.1	824	15	US-10-171-384-1	Sequence 1, Appl
31	94	16.1	840	11	US-09-884-465A-7	Sequence 7, Appl
32	94	16.1	999	11	US-09-884-465A-376	Sequence 376, Appl
33	94	16.1	999	11	US-09-884-465A-377	Sequence 377, Appl
34	94	16.1	1126	11	US-09-884-465A-383	Sequence 383, Appl
35	94	16.1	1238	11	US-09-884-465A-381	Sequence 381, Appl
36	94	16.1	1365	11	US-09-884-465A-382	Sequence 382, Appl
37	92	15.8	572	15	US-10-205-823-225	Sequence 225, Appl
38	91	15.6	52	10	US-09-855-754-17	Sequence 17, Appl
39	91	15.6	56	10	US-09-855-754-18	Sequence 18, Appl
40	91	15.6	439	11	US-09-056-019-37	Sequence 37, Appl
41	90	15.4	52	10	US-09-855-754-21	Sequence 21, Appl
42	90	15.4	324	9	US-09-815-242-11345	Sequence 11345, A
43	89.5	15.3	535	11	US-09-927-827-50	Sequence 50, Appl
44	89	15.2	87	15	US-10-215-432-27	Sequence 27, Appl
45	89	15.2	204	10	US-09-252-150-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-09-919-603-5
Sequence 5, Application US/09919603
Patent No. US20020137679A1
GENERAL INFORMATION:
APPLICANT: Lawler, John W.
TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
FILE REFERENCE: 1440.1033-007
CURRENT APPLICATION NUMBER: US/09/919,603
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: PCT/US00/02482
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/118,053
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 242
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chimeric protein
US-09-919-603-5
```

```
Query Match 44.9%; Score 262.5; DB 10; Length 242;
Best Local Similarity 69.5%; Pred. No. 3.6e-18;
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;
QY 8 LLAALAAVCSAAKSSIGGDCSDLGQMLREIQTALALQVDVRLQGVREITFLKNT 67
DB 10 LTLTALGASGGGSPG---SDLGQMLREIQTALALQVDVRLQGVREITFLKNT 65
QY 68 VMECDACGPO-----PKPKP 82
DB 66 VMECDACGMOOSVRTGLPSVP 87
```

```

RESULT 2
US-09-919-603-7
: Sequence 7, Application US/09919603
: Patent No. US2002015679A1
: GENERAL INFORMATION:
: APPLICANT: Lawler, John W.
: TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
: TITLE OF INVENTION: Chimeric Proteins
: FILE REFERENCE: 1440,1033-007
: CURRENT APPLICATION NUMBER: US/09/919,603
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: PCT/US00/02482
: PRIOR FILING DATE: 2000-02-01
: PRIOR APPLICATION NUMBER: 60/118,053
: PRIOR FILING DATE: 1999-02-01
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: FastSeq For Windows Version 4.0
: SEQ ID NO 7
: LENGTH: 300
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chimeric protein
: US-09-919-603-7

```

Query Match	44.9%	Score 262.5	DB 10	Length 300
Best Local Similarity	69.5%	Pred. No. 4,6	Indels 18	
Matches	57	Conservative	5	Mismatches 9
				Indels 11
				Gaps 2
QY	8	LLALAAVSAAKKGSLLGGDCSDLDGPMQLRLQSTNNAALQDVRWMLRQVREITFLKNT	67	
Db	10	LLTTLAAALASGGGSPFG---SDLGPMQLRLQSTNNAALQDVRWMLRQVREITFLKNT	65	
QY	68	VMECDACGPGQ-----PQPKP	82	
Db	66	VMECDACGMOQSVRTGLPSVRP	87	

```

RESULT 3
US-09-919-603-3
; Sequence 3, Application US/09919603
; Patent No. US20020137679A1
; GENERAL INFORMATION:
; APPLICANT: Lawler, John W.
; TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
; TITLE OF INVENTION: Chimeric Proteins
; FILE REFERENCE: 1440.1033-007
; CURRENT APPLICATION NUMBER: US/09/919,603
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: PCT/US00/02482
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/118,053
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-919-603-3

```

Query Match	44.9%	Score 262.5	DB 10	Length 757
Best Local Similarity	69.5%	Pred. No. 1,3e-17		
Matches	57	Conservative	5	Mismatches 9; Indels 11; Gaps 2
QY	8	LLAALAVSAAKKGSILGDCDSDLGPGMRLQETNALQDVRWMLAQQVREITFLKNT	67	
DB	10	LLTLLAALASGGSGSPLG---SDLGPGMRLQETNALQDVRWMLAQQVREITFLKNT	65	
QY	68	VMECDACGPQ-----DQPKP	82	
DB	66	VMECDACGMQGSVRTGLPSVRP	87	

```

1      RESULT 4
2      US-10-301-822-41
3      ; Sequence 41, Application US/10301822
4      ; Publication No. US20030148410A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Millennium Pharmaceuticals, Inc.
7      ; APPLICANT: Berger, Allison
8      ; APPLICANT: Guillemette, Tracy L.
9      ; APPLICANT: Kamatkar, Shubhangi
10     ; APPLICANT: Schlegel, Robert
11     ; APPLICANT: Monahan, John E.
12     ; APPLICANT: Thibodeau, Stephen N.
13     ; APPLICANT: Burgart, Lawrence J.
14     ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
15     ; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
16     ; FILE REFERENCE: MEM01-029P2RNM
17     ; CURRENT APPLICATION NUMBER: US/10/301,822
18     ; CURRENT FILING DATE: 2002-11-21
19     ; PRIOR APPLICATION NUMBER: US 60/339,971
20     ; PRIOR FILING DATE: 2001-12-10
21     ; PRIOR APPLICATION NUMBER: US 60/361,978
22     ; PRIOR FILING DATE: 2002-03-05
23     ; PRIOR APPLICATION NUMBER: US 60/381,988
24     ; PRIOR FILING DATE: 2002-05-20
25     ; NUMBER OF SEQ ID NOS: 228
26     ; SOFTWARE: FastSeq for Windows Version 4.0
27     ; SEQ ID NO 41
28     ; LENGTH: 757
29     ; TYPE: PRT
30     ; ORGANISM: Homo Sapiens
31     ; US-10-301-822-41

```

Query Match	44.9%	Score 262.5	DB 12	Length 757
Best Local Similarity	69.5%	Pred. No. 1,3e-17		
Matches	57	Conservative	5	Mismatches 9, Indels 11, Gaps 2

OY	8	LLALAANCASAKKSSLGSDCCSPDGSMRELOETNAALQVDPDWKROQVREITLKNT	67
		:::	
Db	10	LLTTLAAGASGGQSPG---SLGPGMRELOETNAALQVDPDWKROQVREITLKNT	65
		:::	
OY	68	VMECDACGPQ-----PQPKP	82
		:::	
Db	66	VMECDACGMQGSVRTGLPSVPR	87
		:::	

```

RESULT 5
US-10-177-293-72
; Sequence 72, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatc, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Ganavayaru, Manjula
; APPLICANT: Kamatkar, Shubhangl
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Baet Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzatzi, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
;

```

US-10-177-293-72

Query Match

Best Local Similarity 44.9%; Score 262.5; DB 15; Length 757;
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

ORGANISM: Homo sapiens

SEQUENCE: FASTSEQ for Windows Version 4.0

SEQUENCE ID NO 72

LENGTH: 757

TYPE: PRT

ORGANISM: Homo sapiens

US-10-177-721-4

Query Match

Best Local Similarity 69.5%; Score 145.5; DB 15; Length 961;
Matches 30; Conservative 10; Mismatches 25; Indels 5; Gaps 2;

ORGANISM: Homo sapiens

SEQUENCE: FASTSEQ for Windows Version 4.0

SEQUENCE ID NO 4

LENGTH: 961

TYPE: PRT

ORGANISM: Homo sapiens

US-10-154-971-22

Query Match

Best Local Similarity 24.0%; Score 140; DB 15; Length 178;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORGANISM: Homo sapiens

SEQUENCE: FASTSEQ for Windows Version 4.0

SEQUENCE ID NO 22

LENGTH: 178

TYPE: PRT

ORGANISM: Homo sapiens

US-10-154-971-24

Query Match

Best Local Similarity 24.0%; Score 140; DB 15; Length 178;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORGANISM: Homo sapiens

SEQUENCE: FASTSEQ for Windows Version 4.0

SEQUENCE ID NO 24

LENGTH: 178

TYPE: PRT

ORGANISM: Homo sapiens

US-10-154-971-22

Query Match

Best Local Similarity 24.0%; Score 140; DB 15; Length 178;
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORGANISM: Homo sapiens

SEQUENCE: FASTSEQ for Windows Version 4.0

SEQUENCE ID NO 22

LENGTH: 178

TYPE: PRT

ORGANISM: Homo sapiens

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/154,971
FILING DATE: 28-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/945,244
FILING DATE: <Unknown>
APPLICATION NUMBER: PCT/EP96/01725
FILING DATE: 25-APR-1996
APPLICATION NUMBER: EP 95400932.0
FILING DATE: 25-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gollin, Michael A.
REGISTRATION NUMBER: 31,957
REFERENCE/DOCKET NUMBER: GUPLA 0003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-414-4000
TELEFAX: 202-414-4040
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-10-154-971-24

Query Match 23.8% Score 139; DB 15; Length 303;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 PGPQPKPQPQPQPQPQPQPQP 98
DB 138 PGPQPKPQPQPQPQPQPQPQP 160

RESULT 9
US-10-156-761-12544
Sequence 12544, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12544
LENGTH: 578
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-12544

Query Match 21.6% Score 126; DB 15; Length 578;
Best Local Similarity 79.2%; Pred. No. 0.00024;
Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 76 PGPQPKPQPQPQPQPQPQPQP 99
DB 36 PGPQPKPQPQPQPQPQPQPQP 59

RESULT 10
US-10-301-822-155
Sequence 155, Application US/10301822
Publication No. US20030148410A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kametkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
APPLICANT: Burgate, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF COLON CANCER
FILE REFERENCE: MPW01-029P2RNM
CURRENT APPLICATION NUMBER: US/10/301,822
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 155
LENGTH: 400
TYPE: PRT
ORGANISM: Homo Sapiens
US-10-301-822-155

Query Match 20.9% Score 122; DB 12; Length 400;
Best Local Similarity 79.2%; Pred. No. 0.0004;
Matches 19; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 76 PGPQPKPQPQPQPQPQPQPQP 99
DB 322 PGPQPKPQPQPQPQPQPQPQP 345

RESULT 11
US-10-280-953-17
Sequence 17, Application US/10280953
Publication No. US20030113317A1
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Arvizu, Chandra S.
APPLICANT: Corley, Neil C.
APPLICANT: Guejler, Karl J.
TITLE OF INVENTION: MOLECULE ASSOCIATED WITH APOPTOSIS
FILE REFERENCE: PC-0018-1 CIP
CURRENT APPLICATION NUMBER: US/10/280,953
CURRENT FILING DATE: 2002-10-23
PRIOR APPLICATION NUMBER: 09/602,565
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/106,920
PRIOR FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PERL Program
SEQ ID NO 17
LENGTH: 261
TYPE: PRT
ORGANISM: Mus musculus
FEATURE: misc.feature
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20030113317A1 g1469400

US-10-280-953-17

Query Match 19.9%; Score 116; DB 15; Length 261;

Best Local Similarity 75.0%; Pred. No. 0.00095;

Matches 18; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 76 PPOPKPQPOPKPQPKPEPE 99
Db 173 PPOPKPQPOPKPQPKPEPE 196

RESULT 12

US-09-881-752A-228

Sequence 228, Application US/09881752A

Patent No. US20020115078A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold

APPLICANT: Al-Garawi, Amal

APPLICANT: Miller, Charles

APPLICANT: Tomb, Jean-Francois

APPLICANT: Oomen, Raymond P.

TITLE OF INVENTION: Identification of Polynucleotides

TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the

FILE REFERENCE: 06132/041002

CURRENT APPLICATION NUMBER: US/09/881,752A

CURRENT FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 08/833,457

PRIOR FILING DATE: 1997-04-01

NUMBER OF SEQ ID NOS: 370

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 228

LENGTH: 285

TYPE: PRT

ORGANISM: Helicobacter pylori

US-09-881-752A-228

Query Match 19.7%; Score 115; DB 10; Length 285;

Best Local Similarity 66.7%; Pred. No. 0.0013;

Matches 16; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 76 PPOPKPQPOPKPQPKPEPE 99
Db 106 PKPKPKPKPKPKPKPKPKPKPK 129

RESULT 13

US-10-138-221-7

Sequence 7, Application US/10138221

Publication No. US20020199216A1

GENERAL INFORMATION:

APPLICANT: Amy F. Macrae

TITLE OF INVENTION: USE OF TRANSPOSABLE ELEMENTS FOR ALTERING GENE EXPRESSION

FILE REFERENCE: 51178/7

CURRENT APPLICATION NUMBER: US/10/138,221

CURRENT FILING DATE: 2002-08-01

PRIOR APPLICATION NUMBER: US 60/287,882

PRIOR FILING DATE: 2001-05-01

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn version 3.1

SEQ ID NO 7

LENGTH: 807

TYPE: PRT

ORGANISM: Zea mays

US-10-138-221-7

Query Match 19.0%; Score 111; DB 14; Length 807;

Best Local Similarity 73.9%; Pred. No. 0.011;

Matches 17; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 77 QPOPKPQPOPKPQPKPEPE 99
Db 108 EPOPKPQPOPKPQPKPEPE 130

RESULT 14

US-09-764-864-967

Sequence 967, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT23

CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 967

LENGTH: 316

TYPE: PRT

ORGANISM: Homo sapiens

US-09-764-864-967

Query Match 18.5%; Score 108; DB 10; Length 316;

Best Local Similarity 70.8%; Pred. No. 0.0072;

Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 76 PPOPKPQPOPKPQPKPEPE 99
Db 231 PPOPKPQPOPKPQPKPEPE 254

RESULT 15

US-09-738-626-6457

Sequence 6457, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIALI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAOKO

APPLICANT: SENO, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 6457

LENGTH: 266

TYPE: PRT

ORGANISM: Corynebacterium glutamicum

US-09-738-626-6457

Query Match 18.4%; Score 107.5; DB 10; Length 266;

Best Local Similarity 29.4%; Pred. No. 0.0067;

Matches 25; Conservative 16; Mismatches 29; Indels 15; Gaps 3;

Qy 39 ELOETNAALQDV---RDW-----LROQVREITFLKXNTVMCDACGPOPO-----PKPO 83
Db 138 QLEEMRTQLSLSLSREKGYEPTMLRAEARRILLESQQLSQOQFOAPQBPVAVPEVPEPM 197
Qy 84 PPOPKPQPOPKPEPECTGSSEKDE 108

Db 198 PEPAPPEPPEPPEVAVAEABEE 222

Search completed: August 26, 2003, 16:07:06
Job time : 39.5841 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 15:54:13 ; Search time 49.1947 Seconds
(without alignments)
213.080 Million cell updates/sec

Title: US-09-696-872-23

Perfect score: 584

Sequence: 1 MRVIMIGLMLAAVCSAAK.....PKQPKPEPGTSSSEKDEL 109

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	234	40.1	755	2 A44315	cartilage oligomer
2	146	25.0	135	2 S33598	Ig gamma-2 chain -
3	145.5	24.9	961	1 TSHUP4	thrombospondin 4 p
4	138.5	23.7	955	2 A45441	thrombospondin 4 -
5	133	22.8	332	1 OZZOMB	circumsporozoite p
6	133	22.8	348	1 OZZOMB	circumsporozoite p
7	132	22.6	3164	1 VMBEHC	U36 protein - hum
8	124	21.2	1494	2 T14355	protein-tyrosine-p
9	122	20.9	400	2 S58222	PO-rich protein -
10	118	20.2	165	2 T24470	hypothetical prote
11	118	20.2	422	2 C91067	hypothetical prote
12	117	20.0	309	2 T17557	procyelin homolog
13	117	20.0	487	2 F83634	hypothetical prote
14	117	20.0	280	2 F71829	sideophore-mediac
15	115	19.7	285	2 E64687	sideophore-mediac
16	115	19.7	378	2 S14959	proline-rich prote
17	113.5	19.4	127	2 S09372	hypothetical prote
18	112.5	19.3	839	1 TQZMCA	probable transpos
19	112.5	19.3	907	2 TQZMCA	hypothetical prote
20	112	19.2	141	2 T09546	extensin like prot
21	112	19.2	283	2 T16343	hypothetical prote
22	111	19.0	115	1 FAUTPC	procyelic acidic r
23	111	19.0	129	2 A26036	procyelic acidic r
24	111	19.0	138	2 B27863	Ac hypothetical pr
25	111	19.0	139	2 S70010	glutamate/proline-
26	111	19.0	143	2 A26918	procyelic acidic r
27	111	19.0	145	2 A44318	procyelin PSSA-1 -
28	111	19.0	221	2 S70009	glutamate/proline-
29	111	19.0	407	2 E82478	hypothetical prote

30	111	19.0	427	2 T03955	probable transpos
31	111	19.0	807	2 T02916	hypothetical prote
32	111	19.0	1749	2 S75138	hypothetical prote
33	110.5	18.9	149	2 B9651	protein T3P18.7 (i
34	110	18.8	792	2 F90566	conserved hypothet
35	108.5	18.6	312	2 B84632	probable chloropla
36	108	18.5	266	2 T44781	tonB protein (impo
37	107	18.3	437	2 T26767	hypothetical prote
38	107	18.3	633	1 S49611	probable serine/th
39	107	18.3	833	2 A90575	conserved hypothet
40	105	18.0	877	2 F90070	clumping factor B
41	104	17.8	949	2 T08658	hypothetical prote
42	104	17.8	1569	2 A65044	hypothetical prote
43	103.5	17.7	649	2 T18063	DNA binding protei
44	102	17.5	430	2 TC2301	hypothetical 47.8k
45	102	17.5	449	2 S16748	proline-rich prote

ALIGNMENTS

RESULT 1
A44315
cartilage oligomeric matrix protein precursor - rat
N:Alternate names: thrombospondin homolog COMP
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Apr-1993 #sequence_revision 03-May-1994 #text_change 02-Aug-2002
C:Accession: A44315
R:Oldberg, A.; Antonsson, P.; Lindblom, K.; Heinegard, D.
J. Biol. Chem. 267, 22346-22350, 1992
A:Title: COMP (cartilage oligomeric matrix protein) is structurally related to the throm
A:Reference number: A44315; MUID:93054522; PMID:1429587
A:Accession: A44315
A:Molecule type: mRNA
A:Residues: 1-755 <OLD>
A:Cross-references: GB:472914; NID:6297438; PIDN:CA51419.1; PID:6297439
A:Experimental source: tracheal chondrocyte
A>Note: sequence extracted from NCBI backbone (NCBI:P117022)
C:Superfamily: thrombospondin 3; EGF homology
C:Keywords: pentamer
F:1-19/Domain: signal sequence #status predicted <SIG>
F:182-219/Domain: EGF homology <EGF>

Query Match 40.1%; Score 234; DB 2; Length 755;
Best local similarity 63.0%; Pred. No. 4.6e-13;
Matches 51; Conservative 8; Mismatches 18; Indels 4; Gaps 1;

Oy 8 LIALAAVCSAAKSSIGDCSDLPQMLRELOETNAALQDVARDMDROQVREITPKNT 67
:|||||: : ||| ||||| ||||| ||||| : : |||||
Db 9 VLIALAARATGGGQIGPVG----DLAPQMLRELOETNAALQDVRELRHVKETPKNT 64

Oy 68 VMECDACGPOPQPKPOPQPOP 88
:|||||: ||||| ||||| ||||| :
Db 65 VMECDACGMPARTPGSLVRP 85

RESULT 2
S33598
Ig gamma-2 chain - Arabian camel (fragment)
C:Species: Camelus dromedarius (Arabian camel)
C>Date: 19-Mar-1997 #sequence_revision 10-Oct-1997 #text_change 11-Jan-2000
R:Hamers-Castelman, C.; Atachouch, T.; Myltermans, S.; Robinson, G.; Songa
Nature 363, 446-448, 1993
A:Title: Naturally occurring antibodies devoid of light chains.
A:Reference number: S33598; MUID:93275410; PMID:8502296
A:Accession: S33598
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-135 <HAM>
A:Experimental source: spleen
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: disulfide bond; immunoglobulin

F:14-62/Disulfide bonds: #status predicted

Query Match 25.0%; Score 146; DB 2; Length 135;
Best Local Similarity 96.2%; Pred. No. 4,8e-06;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 76 PQPQPKPQPQPQPKPQPKPEEGT 101
DB 91 PQPQPKPQPQPQPKPQPKPEECT 116

RESULT 3

75HUP4

thrombospondin 4 precursor - human

C:Species: Homo sapiens (man)

C:Date: 06-Jan-1995 #sequence_revision 11-Aug-1995 #text_change 13-Aug-1999

C:Accession: A55710; S36069

R:Lawler, J.; McHenry, K.; Duquette, M.; Derrick, L.

U. Biol. Chem. 270, 2809-2814, 1995

A:Title: Characterization of human thrombospondin-4.

A:Reference number: A55710; MUID:95155352; PMID:7852353

A:Accession: A55710

A:Molecule type: mRNA

A:Residues: 1-961 <LAN>

A:Cross-references: EMBL:Z19585; NID:g311625; PIDN:CAA79635.1; PID:g311626

A>Note: authors translated the codon GTG for residue 616 as Ser

C:Genetics:

A:Gene: GDB: T884

A:Cross-references: GDB:463011; OMIM:600715

A:Map position: 1q21-q23

C:Complex: homotrimer, disulfide linked

C:Function: A:Description: participates in cell migration and adhesion, and in platelet aggregation

C:Superfamily: thrombospondin 3; EGF homology

C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; cell adh

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-961/Product: thrombospondin 4 #status predicted <MAT>

F:290-324/Domain: EGF homology <EGF1>

F:330-362/Domain: EGF homology <EGF2>

F:562-564/Region: cell attachment (R-G-D) motif

F:303/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted

F:343/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

F:612/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.9%; Score 145.5; DB 1; Length 961;
Best Local Similarity 42.9%; Pred. No. 3,6e-05;
Matches 30; Conservative 10; Mismatches 25; Indels 5; Gaps 2;

QY 31 DLGPMRLBELQETNALQDVRWLRQVREITFLKNTVMECDACGP---QPQPKPQPK 86
DB 218 DENRQFLQMTQNLQGLGVKDLRQVETSFLLNTIAECQCGPLKFGSPPTVVAP 277

QY 87 QPQPKPQPK 96
DB 278 AP-PAPPRP 286

RESULT 4

A45441

thrombospondin 4 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000

C:Accession: A45441

R:Lawler, J.; Duquette, M.; Whitaker, C.A.; Adams, J.C.; McHenry, K.; Deslmonne, D.W.

J. Cell Biol. 120, 1059-1067, 1993

A:Title: Identification and characterization of thrombospondin-4, a new member of the th

A:Reference number: A45441; MUID:93163109; PMID:8432726

A:Accession: A45441

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-955 <LAN>

A:Cross-references: GB:Z19091; NID:g288777; PIDN:CAA79518.1; PID:g288778

A>Note: sequence extracted from NCBI backbone (NCBIN:124858, NCBI:124860)

C:Superfamily: thrombospondin 3; EGF homology
F:285-319/Domain: EGF homology <EGF>

Query Match 23.7%; Score 138.5; DB 2; Length 955;
Best Local Similarity 34.3%; Pred. No. 0.00015;
Matches 37; Conservative 13; Mismatches 31; Indels 27; Gaps 4;

QY 23 SLGDCSDLPQMLBELQETNALQDVRWLRQVREITFLKNTVMECDACGPQ-PK 81
DB 208 SEAGQGTQDVSRQRLIGQTQNNQMLGELRDVWRQVETMFLLNTIAECQCGPPEPL 267

QY 82 PQPQPKPQPKPQPKPQPKPEEGT-TGS 103
DB 268 PTKVFORLATTTPPKPRCDATSCFRGVRCIDTEGFGQPCPEEGTGN 315

RESULT 5

OZZQNB

circumsporozoite protein precursor - Plasmodium berghei (strain NK65)

N:Alternate names: sporozoite surface antigen

C:Species: Plasmodium berghei

C:Date: 30-Sep-1987 #sequence_revision 28-Jul-1995 #text_change 16-Jul-1999

C:Accession: A44948; A25083; S13446

R:Janar, D.E.

Mol. Biochem. Parasitol. 39, 151-154, 1990

A:Title: Sequence of the circumsporozoite gene of Plasmodium berghei ANKA clone and NK65

A:Reference number: A44948; MUID:90158693; PMID:2406593

A:Accession: A44948

A:Molecule type: DNA

A:Residues: 1-332 <LAN>

A:Cross-references: GB:M2887

R:Elchinger, D.J.; Arnott, D.E.; Tam, J.P.; Nussenzweig, V.; Bnea, V.

Mol. Cell. Biol. 6, 3965-3972, 1986

A:Title: Circumsporozoite protein of Plasmodium berghei: gene cloning and identification

A:Reference number: A25083; MUID:87089740; PMID:2432395

A:Accession: A25083

A:Molecule type: DNA

A:Residues: 1-26, 'I', 28-68, 'PMLRR', 75-126, 'P', 128-134, 'PPNANDP', 135-332 <EIC>

A:Cross-references: GB:M4135; NID:g160245; PIDN:AAA29577.1; PID:g160246

R:Weber, J.L.; Egan, J.E.; Lyon, J.A.; Wirtz, R.A.; Charonvlt, Y.; Maloy, W.L.; Hockmeyer

Exp. Parasitol. 63, 295-300, 1987

A:Title: Plasmodium berghei: cloning of the circumsporozoite protein gene.

A:Reference number: S13446; MUID:87218962; PMID:3556207

A:Accession: S13446

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 61-122, 'A', 124-332 <WEB>

A:Cross-references: GB:M25445; NID:g160177; PIDN:AAA29531.1; PID:g160178

C:Comment: There are three distinct regions in the mature circumsporozoite protein, the e

obic membrane-anchoring sequence.

C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

C:Keywords: sporozoite; surface antigen; tandem repeat

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-332/Product: circumsporozoite protein #status predicted <MAT>

F:194-189/Region: 8-residue repeats

F:199-230/Region: 2-residue repeats

F:258-310/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 22.8%; Score 133; DB 1; Length 332;
Best Local Similarity 62.5%; Pred. No. 0.00016;
Matches 20; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 76 PQPQPKPQPQPQPKPQPKPEEGTSSKSD 107
DB 209 PQPQPKPQPQPQPKPQPKPEEGTNNNNKN 240

RESULT 6

OZZQNB

circumsporozoite protein precursor - Plasmodium berghei (strain ANKA clone 2.34L)

N:Alternate names: sporozoite surface antigen

C:Species: Plasmodium berghei

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

CJAccession: S07873; S12571
R.Lockyer, M.J.; Davies, C.S.; Sunbrier, A.; Sinden, R.E.
Nucleic Acids Res. 18, 376, 1990
A>Title: Nucleotide sequence of the Plasmidium berghei circumsporozoite protein gene from
A:Reference number: S07873; MUID:90221834; PMID:2183186
A:Accession: S07873
A:Molecule type: DNA
A:Residues: 1-348 <LOC>
A:Cross-references: EMBL:X17606
R.Lockyer, M.J.
submitted to the EMBL Data Library, November 1989
A:Reference number: S12571
A:Accession: S12571
A:Molecule type: DNA
A:Residues: 1-59,'T',61-81,83-348 <LOC2>
A:Cross-references: EMBL:X17606; NID:g9784; PID:CAA35608.1; PID:g9785
C:Superfamily: Circumsporozoite protein; thrombospondin type 1 repeat homology
C:Keywords: tandem repeat
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-348/Product: circumsporozoite protein #status predicted <MAR>
F:94-205/Region: 8-residue repeats
F:215-247/Region: 2-residue repeats
F:274-326/Domain: thrombospondin type 1 repeat homology <THRI>

Query Match 22.8% Score 133; DB 1; Length 348;
Best Local Similarity 62.5%; Pred. No. 0.00017;
Matches 20; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Oy 76 P Q P K P Q P O P Q P K P P K P E P E G T G S E K D 107
 |||||:|||||:|||||:|||||:::|
Db 225 P Q P O P Q P O P Q P O P R P Q P O P G C N N N K X N 256

RESULT 7
WMBEH6
UJ36 proteain - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000
C:Accession: J30085
R.McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perrin,
J. Gen. Virol. 69, 1531-1574, 1988
A>Title: The complete DNA sequence of the long unique region in the genome of herpes sim
A:Reference number: A30083; MUID:88274327; PMID:2839594
A:Accession: J30085
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3164 <MEG>
A:Cross-references: GB:X14112; NID:g1944536; PID:N:CAA3311.1; PID:g59536; GB:D00317
C:Genetics:
A:Gene: UJ36
C:Superfamily: varicella-zoster virus gene 22 protein

Query Match 22.6% Score 132; DB 1; Length 3164;
Best Local Similarity 66.7%; Pred. No. 0.0017;
Matches 22; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 67 T Y M E C D A C G P O P Q P O P Q P O P Q P K P K P E P E 99
 |||||:|||||:|||||:|||||:|||||
Db 2904 T V S R L S A P O P Q P O P Q P O P Q P O P Q P O P Q P Q 2936

RESULT 8
T14355
protein-tyrosine-phosphatase (EC 3.1.3.48) TD14 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14355
R.Cao, L.; Zhang, L.; Ruiz-Lozano, P.; Yang, Q.; Chien, K.R.; Graham, R.M.; Zhou, M.
J. Biol. Chem. 273, 21077-21083, 1998
A>Title: A novel putative protein-tyrosine phosphatase contains a BROL-like domain and s
A:Reference number: Z18004; MUID:98361981; PMID:9694860
A:Accession: T14355
A>Status: preliminary; translated from GB/EMBL/DBD

```

A:Molecule type: mRNA
A:Residues: 1-1494 <CNO>
A:Cross-references: EMBL:AF077000; NID:g3598973; PID:g3598974; PIDN:AAC62959.1
A:Experimental source: brain
C:Genetics:
A:Gene: PTP-TD14
C:Function:
A:Description: may be involved in regulating Ha-ras-dependent cell growth
C:Keywords: phosphoric monoester hydrolase

Query Match      21.2%; Score 124; DB 2; Length 1494;
Best Local Similarity 79.2%; Pred. No. 0.0043;
Matches 19; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY      76 PQQPQPQPQPQPQPQPQPQPQP 99
DB      797 PQQPQPQPQPQPQPQPQPQPQP 820

RESULT 9
PQ-rich protein - human
S58222
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #ext_change 05-Nov-1999
C:Accession: S58222
R:Wagner, F.F.; Plegel, W.A.
Submitted to the EMBL Data Library, July 1995
A:Description: A cDNA, which predicts a protein with PQ-rich repeats, isolated from a phe
A:Reference number: S58222
A:Accession: S58222
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-400 <MAC>
A:Cross-references: EMBL:Z50194; NID:g929659; PIDN:CAA90576.1; PID:g929660

Query Match      20.9%; Score 122; DB 2; Length 400;
Best Local Similarity 79.2%; Pred. No. 0.0018;
Matches 19; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY      76 PQQPQPQPQPQPQPQPQPQPQP 99
DB      322 PQQPQPQPQPQPQPQPQPQPQP 345

RESULT 10
T24470
hypochemical protein T04F8.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 18-Feb-2000
C:Accession: T24470
R:Lennard, N.
Submitted to the EMBL Data Library, November 1995
A:Reference number: Z19895
A:Accession: T24470
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-165 <WLL>
A:Cross-references: EMBL:Z66565; PIDN:CAA91483.1; GSPDB:GN00028; CESP:T04F8.8
A:Experimental source: clone T04F8
C:Genetics:
A:Gene: CESP:T04F8.8
A:Map position: X
A:Introns: 18/2; 63/1

Query Match      20.2%; Score 118; DB 2; Length 165;
Best Local Similarity 66.7%; Pred. No. 0.0017;
Matches 16; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

OY      76 PQQPQPQPQPQPQPQPQPQPQP 99
DB      108 PEPBPBPBPBPBPBPBPBPBPBP 131

```


This Page Blank (unspooled)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:01:13 ; Search time 27.0089 Seconds
(without alignments)
189.786 Million cell updates/sec

Title: US-09-696-872-23
Perfect score: 584
Sequence: 1 MRVWILGLALAAVCSAAK.....PKPQKPEPEGTGSSEKDEL 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262.5	44.9	757	1 COMP_HUMAN	P49747 homo sapien
2	234	40.1	755	1 COMP_RAT	P35444 rattus norv
3	145.5	24.9	961	1 TSP4_HUMAN	P35443 homo sapien
4	144.5	24.7	980	1 TSP4_RAT	P49744 rattus norv
5	138.5	23.7	955	1 TSP4_XENLA	O06441 xenopus lae
6	133	22.8	339	1 CSP_PLAAB	P06915 plasmodium
7	133	22.8	347	1 CSP_PLAAB	P23093 plasmodium
8	132	22.6	3164	1 TEGU_HSV11	P10220 herpes simp
9	117	20.0	220	1 NOL3_MOUSE	O941X0 mus musculu
10	115	19.7	280	1 TONB_HELPJ	O941X0 mus musculu
11	115	19.7	285	1 TONB_HELPJ	O28899 helicobacte
12	112.5	19.3	839	1 TRAG_MAIZE	P03010 zea mays (m
13	111	19.0	115	1 PARB_TRYBB	P14043 trypanosoma
14	111	19.0	129	1 PARB_TRYBB	P09791 trypanosoma
15	111	19.0	143	1 PARC_TRYBB	P08469 trypanosoma
16	111	19.0	145	1 PARC_TRYBB	O06084 trypanosoma
17	111	19.0	221	1 NOL3_RAT	O06881 rattus norv
18	111	19.0	806	1 TRAI_MAIZE	P08770 zea mays (m
19	108	18.5	722	1 Z219_HUMAN	O9P2Y4 homo sapien
20	107	18.3	613	1 PKPA_PHYL	O01577 phycomyces
21	104	17.8	1569	1 YPJA_ECOLI	P52143 escherichia
22	102	17.5	449	1 ARG_BRANA	P40603 brassica na
23	101	17.3	342	1 TONB_PSEAE	O51368 pseudomonas
24	101	17.3	450	1 V3OK_BYDVP	P09516 barley yell
25	101	17.3	565	1 TSP3_MOUSE	O70324 mus musculu
26	100.5	17.2	956	1 TSP3_MOUSE	P49746 homo sapien
27	100	17.1	643	1 VP40_HSV2	P52369 equine herp
28	100	17.1	956	1 TSP3_MOUSE	O05895 mus musculu
29	99	17.0	1493	1 M3K1_RAT	O62925 rattus norv
30	99	17.0	1520	1 ACPD_ECOLI	O48837 escherichia
31	98	16.8	534	1 AFG_AATH	P46602 arabidopsis
32	97	16.6	243	1 TONB_KLEBN	P46610 klebsiella
33	96.5	16.5	737	1 CTCF_RAT	O91d1 rattus norv

34	96	16.4	613	1 MOT8_HUMAN	P36021 homo sapien
35	96	16.4	1850	1 BAZA_MOUSE	O91Y65 mus musculu
36	95.5	16.4	605	1 GLCA_SOYBN	P13916 glycine max
37	95	16.3	159	1 E3GL_ADE02	P03251 human adeno
38	94.5	16.2	243	1 TONB_ENTAE	P46383 enterobacte
39	94.5	16.2	411	1 ALX4_HUMAN	O9161 homo sapien
40	93.5	16.0	525	1 LAG3_HUMAN	P18627 homo sapien
41	93	15.9	1250	1 YFAL_ECOLI	P45508 escherichia
42	92	15.8	572	1 LMD1_HUMAN	P29536 homo sapien
43	92	15.8	847	1 Y083_NPYAC	O06570 autographa
44	91.5	15.7	2004	1 MOZ_HUMAN	O92794 homo sapien
45	91	15.6	791	1 Y046_UREPA	O9199 ureaplasma

ALIGNMENTS

RESULT 1
COMP_HUMAN STANDARD: PRT; 757 AA.
AC P49747; O16388; O16389;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cartilage oligomeric matrix protein precursor (COMP).
GN COMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage; PubMed=7713493.
RX MEDLINE=95229140; PubMed=7713493.
RA Newton G., Weremowicz S., Morton C.C., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Lawler J.;
RT "Characterization of human and mouse cartilage oligomeric matrix
RT protein";
RL Genomics 24:435-439(1994).
RN [2]
RP VARIANTS PSACH SER-459 DEL; TYR-468 AND TYR-472.
RX MEDLINE=95400301; PubMed=7670471.
RA Hecht J.T., Nelson L.D., Crowder E., Wang Y., Elder F.F.B.,
RA Harrison W.R., Francosano C.A., Prange C.K., Lennon G.G., Deere M.,
RA Lawler J.;
RT "Mutations in exon 17B of cartilage oligomeric matrix protein (COMP)
RT cause pseudoachondroplasia";
RL Nat. Genet. 10:325-329(1995).
RN [3]
RP VARIANTS MED TYR-342, AND VARIANT PSACH ARG-328.
RX MEDLINE=95400302; PubMed=7670472.
RA Briggs M.D., Hoffman S.W.G., King L.M., Olsen A.S., Mohrenweiser H.,
RA Leroy J.G., Mortier G.R., Rimoin D.L., Lachman R.S., Gaines E.S.,
RA Cederblad J.A., Knowlton R.G., Cohn D.H.;
RT "Pseudoachondroplasia and multiple epiphyseal dysplasia due to
RT mutations in the cartilage oligomeric matrix protein gene";
RL Nat. Genet. 10:330-336(1995).
RN [4]
RP VARIANT MED LYS-523.
RX MEDLINE=97171141; PubMed=9021009;
RA Ballo R., Briggs M.D., Cohn D.H., Knowlton R.G., Beighton P.H.,
RA Ramegar R.S.;
RT "Multiple epiphyseal dysplasia, ribbing type: a novel point mutation
RT in the COMP gene in a South African family";
RL Am. J. Med. Genet. 68:396-400(1997).
RN [5]
RP VARIANT MED SER-371, AND VARIANT PSACH 513-VAL- LYS-516 DEL.
RX MEDLINE=97327574; PubMed=9184241.
RA Susic S., McGroarty J., Ahler J., Cole W.G.;
RT "Multiple epiphyseal dysplasia and pseudoachondroplasia due to novel
RT mutations in the calmodulin-like repeats of cartilage oligomeric
RT matrix protein";
RL Clin. Genet. 51:219-224(1997).

RN [6] VARIANTS PSACH AND MED.
 RX MEDLINE=98130533; PubMed=9463320;
 RA Briggance M.D., Mortier G.R., Cole W.G., King L.M., Golik S.S.,
 RA Bonaventura V.D., Nuytjck L., de Paeppe A., Leroy J.G., Biesecker L.,
 RA Lipson M., Wilcox W.R., Lachman R.S., Rimoin D.L., Knowlton R.G.,
 RA Cohn D.H.;
 RT "diverse mutations in the gene for cartilage oligomeric matrix protein
 RT in the pseudochondroplasia-multiple epiphyseal dysplasia disease
 RT spectrum.";
 RL Am. J. Hum. Genet. 62:311-319(1998).
 RN [7]
 RN VARIANTS PSACH AND MED.
 RX MEDLINE=99118868; PubMed=9921895;
 RA Ikegawa S., Ohashi H., Nishimura G., Kim K.C., Samnolte A.,
 RA Kimizuka M., Fukushima Y., Nagai T., Nakamura Y.;
 RA "Novel and recurrent COMP (cartilage oligomeric matrix protein)
 RT mutations in pseudochondroplasia and multiple epiphyseal dysplasia.";
 RL Hum. Genet. 103:633-638(1998).
 RN [8]
 RN VARIANTS PSACH AND MED.
 RX MEDLINE=98112405; PubMed=9452026;
 RA Loughlin J., Ireen C.H., Muecata Z., Briggs M.D., Carr A., Lynch S.-A.,
 RA Knowlton R.G., Cohn D.H., Sykes B.;
 RT "Identification of five novel mutations in cartilage oligomeric
 RT matrix protein gene in pseudochondroplasia and multiple epiphyseal
 RT dysplasia.";
 RL Hum. Mutat. Suppl. 1:S10-S17(1998).
 RN [9]
 RN VARIANT PSACH GLY-482.
 RX MEDLINE=98112442; PubMed=9452063;
 RA Susic S., Ahler J., Cole W.G.;
 RT "Pseudochondroplasia due to the substitution of the highly conserved
 RT Asp482 by Gly in the seventh calmodulin-like repeat of cartilage
 RT oligomeric matrix protein.";
 RL Hum. Mutat. Suppl. 1:S125-S127(1998).
 CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.
 CC -1- DISEASE: DEFECTS IN COMP ARE THE CAUSE OF PSEUDOCHONDROPLASIA
 CC (PSACH) AND MULTIPLE EPIPHYSEAL DYSPLASIA (MED OR EDM1), WHICH ARE
 CC DOMINANTLY INHERITED CHONDRODYSPLASIAS CHARACTERIZED BY SHORT
 CC STATURE AND EARLY-ONSET OSTEOARTHRITIS. MED IS BROADLY CATEGORIZED
 CC INTO THE MORE SEVERE FAIRBANK AND THE Milder RIBBING TYPES. PSACH
 CC IS MORE SEVERE AND IS RECOGNIZED IN EARLY CHILDHOOD.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: Contains 4 EGF-like domains.
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license@ebi.ch (see <http://www.ebi.ac.uk/announce/>)
 CC or send an email to license@ebi.ac.uk.
 CC -----
 DR EMBL, I32137; AAA57253.1; -;
 DR EMBL, S79499; AAB35269.1; -;
 DR EMBL, S79500; AAB35270.1; -;
 DR HSSP, P35444; 1VDF.
 DR Genew, HGNC:2227; COMP.
 DR MIM, 600310; -;
 DR MIM, 132400; -;
 DR MIM, 177170; -;
 DR GO, GO:0005578; C:extracellular matrix; TAS.
 DR GO, GO:0005509; F:calcium ion binding activity; TAS.
 DR GO, GO:0005201; F:extracellular matrix structural constituent; TAS.
 DR GO, GO:0007397; P:histogenesis and organogenesis; TAS.
 DR GO, GO:0001501; P:skeletal development; TAS.
 DR InterPro, IPR001681; EGF_Ca.
 DR InterPro, IPR006209; EGF_Like.
 DR InterPro, IPR003367; tep_3.
 DR Pfam, PF00008; EGF_2.
 DR Pfam, PF02412; tep_3; 11.

DR	SMART; SMO0179; EGF CA; 2.			POTENTIAL.
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.			CARTILAGE OLIGOMERIC MATRIX PROTEIN.
DR	PROSITE; PS01186; EGF_2; 1.			N-TERMINAL.
KM	Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain			
KM	Signal; Disease mutation.			
FT	SIGNAL	1	20	
FT	CHAIN	21	757	
FT	DOMAIN	22	86	
FT	DOMAIN	87	126	
FT	DOMAIN	127	179	
FT	DOMAIN	180	222	
FT	DOMAIN	225	267	
FT	DOMAIN	297	332	
FT	DOMAIN	333	355	
FT	DOMAIN	356	391	
FT	DOMAIN	392	414	
FT	DOMAIN	415	452	
FT	DOMAIN	453	488	
FT	DOMAIN	489	524	
FT	DOMAIN	525	757	
FT	DISUPEID	69	69	
FT	DISUPEID	72	72	
FT	DISUPEID	91	102	
FT	DISUPEID	96	111	
FT	DISUPEID	114	125	
FT	DISUPEID	131	142	
FT	DISUPEID	136	151	
FT	DISUPEID	154	178	
FT	DISUPEID	184	197	
FT	DISUPEID	191	206	
FT	DISUPEID	209	221	
FT	DISUPEID	229	243	
FT	DISUPEID	237	253	
FT	DISUPEID	255	266	
FT	DISUPEID	282	287	
FT	DISUPEID	292	312	
FT	DISUPEID	328	348	
FT	DISUPEID	351	371	
FT	DISUPEID	387	407	
FT	DISUPEID	410	430	
FT	DISUPEID	448	468	
FT	DISUPEID	484	504	
FT	DISUPEID	520	741	
FT	CARBOHYD	121	121	
FT	CARBOHYD	742	742	
FT	CARBOHYD	367	369	
FT	SITE	290	290	
FT	VARIANT			
FT	VARIANT	299	299	
FT	VARIANT	328	328	
FT	VARIANT	342	342	
FT	VARIANT	349	349	
FT	VARIANT	361	361	
FT	VARIANT	361	361	
FT	VARIANT	367	368	
FT	VARIANT	371	371	
FT	VARIANT	372	372	
FT	VARIANT	374	374	
FT	VARIANT	387	387	
FT	VARIANT	391	394	

Query Match 44.9%; Score 262.5; DB 1; Length 757;
 Best Local Similarity 69.5%; Pred. No. 2.7e-15;
 Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

Db 8 L1ALAACSAAKSSIGDCCSDLPQMLRELOETNALQDVVDWLRQVREITFLKNT 67
 10 L1TIALGASGGSPG---SDLGPMLELOETNALQDVVDWLRQVREITFLKNT 65

QY 68 VMECDACGPO-----PQPKP 82
 66 VMECDACGMOQSVRTGLPSVRP 87

RESULT 2
 COMP_RAT STANDARD; PRT; 755 AA.

AC P35444;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cartilage oligomeric matrix protein precursor (COMP).
 GN COMP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Cartilage;
 RA MEDLINE=93054522; PubMed=1429587;
 RX Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;
 RT "COMP (cartilage oligomeric matrix protein) is structurally related
 to the thrombospondins.";
 RL J. Biol. Chem. 267:22346-22350(1992).
 RN [2]
 X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 27-72.
 RX MEDLINE=97020114; PubMed=8864111;
 RA Malashkevich V.N., Kammerer R.A., Efimov V.P., Schultness T.,
 RA Engel J.;
 RT "The crystal structure of a five-stranded coiled coil in COMP: a
 prototype ion channel?";
 RL Science 274:761-765(1996).
 CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: Contains 4 EGF-like domains.
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X72914; CAAS1419.1; -
 CC PIR; A44315; A44315.
 CC PDB; 1VDF; 08-OCT-97.
 CC PDB; 1FBM; 09-AUG-00.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_Like.
 CC InterPro; IPR003367; TSP_3.
 CC Pfam; PF00008; EGF; 1.
 CC Pfam; PF02412; TSP_3; 9.
 CC SMART; SM00179; EGF_CA; 2.
 CC PROSITE; PS00022; EGF_1; FALSE_NEG.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS01187; EGF_CA; 2.
 CC Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
 KW Signal; 3D-structure.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 755 CARTILAGE OLIGOMERIC MATRIX PROTEIN.

FT DOMAIN. 21 84 N-TERMINAL.
 FT DOMAIN 85 124 EGF-LIKE 1.
 FT DOMAIN 125 177 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 178 220 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 223 265 EGF-LIKE 4.
 FT DOMAIN 295 330 TSP TYPE-3 1.
 FT DOMAIN 331 353 TSP TYPE-3 2.
 FT DOMAIN 354 389 TSP TYPE-3 3.
 FT DOMAIN 390 412 TSP TYPE-3 4.
 FT DOMAIN 413 450 TSP TYPE-3 5.
 FT DOMAIN 451 486 TSP TYPE-3 6.
 FT DOMAIN 487 522 TSP TYPE-3 7.
 FT DOMAIN 523 755 C-TERMINAL.
 FT DISULFID 68 68 INTERCHAIN.
 FT DISULFID 71 71 INTERCHAIN.
 FT DISULFID 89 100 BY SIMILARITY.
 FT DISULFID 94 109 BY SIMILARITY.
 FT DISULFID 112 123 BY SIMILARITY.
 FT DISULFID 129 140 BY SIMILARITY.
 FT DISULFID 134 149 BY SIMILARITY.
 FT DISULFID 152 176 BY SIMILARITY.
 FT DISULFID 182 195 BY SIMILARITY.
 FT DISULFID 189 204 BY SIMILARITY.
 FT DISULFID 207 219 BY SIMILARITY.
 FT DISULFID 227 241 BY SIMILARITY.
 FT DISULFID 235 251 BY SIMILARITY.
 FT DISULFID 253 264 BY SIMILARITY.
 FT DISULFID 280 285 BY SIMILARITY.
 FT DISULFID 290 310 BY SIMILARITY.
 FT DISULFID 326 346 BY SIMILARITY.
 FT DISULFID 349 369 BY SIMILARITY.
 FT DISULFID 385 405 BY SIMILARITY.
 FT DISULFID 408 428 BY SIMILARITY.
 FT DISULFID 446 466 BY SIMILARITY.
 FT DISULFID 482 502 BY SIMILARITY.
 FT DISULFID 518 739 BY SIMILARITY.
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 740 740 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT HELIX 30 66
 FT TURN 67 67
 FT HELIX 69 71
 FT HELIX 71 71
 SQ SEQUENCE 755 AA; 82663 MW; AB48888FE093C598 CRC64;

Query Match 40.1%; Score 234; DB 1; Length 755;
 Best Local Similarity 63.0%; Pred. No. 7.3e-13;
 Matches 51; Conservative 8; Mismatches 18; Indels 4; Gaps 1;

QY 8 L1ALAACSAAKSSIGDCCSDLPQMLRELOETNALQDVVDWLRQVREITFLKNT 67
 9 VLALALRATCGGIPPLGG---DLAPQMLELOETNALQDVVDWLRQVREITFLKNT 64

Db 68 VMECDACGPOPKPQPOP 88
 65 VMECDACGMOQSVRTGLPSVRP 85

RESULT 3
 ID TSP4 HUMAN STANDARD; PRT; 961 AA.
 AC P35443;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Thrombospondin 4 precursor.
 GN THBS4 OR TSP4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=93353522; PubMed=8350346;

DR Pfam; PF02210; TSPN; 1.
 DR SMART; SM00179; EGF CA; 2.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF CA; 2.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
 KW Signal.

FT SIGNAL 1 39
 FT CHAIN 40 980
 FT DOMAIN 40 303
 FT DOMAIN 304 343
 FT DOMAIN 344 396
 FT DOMAIN 397 437
 FT DOMAIN 438 481
 FT DOMAIN 511 546
 FT DOMAIN 547 569
 FT DOMAIN 570 605
 FT DOMAIN 606 629
 FT DOMAIN 629 666
 FT DOMAIN 667 706
 FT DOMAIN 707 742
 FT DOMAIN 743 980

FT DISULFID 276 279
 FT DISULFID 279 279
 FT DISULFID 308 319
 FT DISULFID 313 328
 FT DISULFID 331 342
 FT DISULFID 348 359
 FT DISULFID 353 368
 FT DISULFID 371 395
 FT DISULFID 401 412
 FT DISULFID 406 421
 FT DISULFID 424 436
 FT DISULFID 442 456
 FT DISULFID 450 466
 FT DISULFID 468 480
 FT DISULFID 496 501
 FT DISULFID 506 526
 FT DISULFID 542 562
 FT DISULFID 565 585
 FT DISULFID 601 621
 FT DISULFID 624 644
 FT DISULFID 662 682
 FT DISULFID 702 722
 FT DISULFID 738 759
 FT CARBOHYD 631 651
 FT CARBOHYD 960 960
 SQ SEQUENCE 980 AA; 108213 MW; 056D41B86206FCF CRC64;

Query Match 24.7%; Score 144.5; DB 1; Length 980;
 Best Local Similarity 42.0%; Pred. No. 4.2e-05;
 Matches 29; Conservative 10; Mismatches 27; Indels 3; Gaps 1;

QY 31 DLGPMRELEQETAAALQDVDMRLQVREITPKNTWEGCDAGPQPPQPPQ---PQ 87
 DB 236 DFNQPFGLQGMQNLQGLGEVLDLRQGVKRSFLNTIAECCAGCPSPFQPTNTLVPI 295

QY 88 PQRPPQPP 96
 DB 296 APPAPPPR 304

RESULT 5
 TSP4_XENLA STANDARD; PRT; 955 AA.
 AC Q06441;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 26-FEB-2003 (Rel. 41, Last annotation update)
 DE Thrombospondin 4 precursor.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93163109; PubMed=8432726;
 RA Lawler J., Duguet M., Whitaker C.A., Adams J.C., McHenry K.,
 RT Desimone D.W.;
 RT "Identification and characterization of thrombospondin-4, a new
 member of the thrombospondin gene family.";
 RL J. Cell Biol. 120:1059-1067 (1993).
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ AND TYPE V COLLAGEN. MAY PARTICIPATE IN THE GENESIS AND
 FUNCTION OF CARDIAC AND SKELETAL MUSCLE.
 CC -1- SUBUNIT: Homotrimer; disulfide-linked.
 CC -1- DEVELOPMENTAL STAGE: INITIAL EXPRESSION DURING NEURULATION.
 CC INCREASE DURING TALBD STAGES BUT DECREASE BY THE FEEDING TADPOLE
 STAGE.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: Contains 4 EGF-like domains.
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z19091; CA79518.1; -.
 DR PIR; A45441; A45441.
 DR HSSP; P35444; IVDF.
 DR InterPro; IPR001881; EGF CA.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR003367; TSP 3.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF02412; TSP 3; 9.
 DR Pfam; PF02210; TSPN; 1.
 DR SMART; SM00179; EGF CA; 2.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF CA; 2.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
 KW Signal.

FT SIGNAL 1 24
 FT CHAIN 25 955
 FT DOMAIN 25 280
 FT DOMAIN 281 320
 FT DOMAIN 321 373
 FT DOMAIN 374 417
 FT DOMAIN 418 459
 FT DOMAIN 489 524
 FT DOMAIN 525 547
 FT DOMAIN 548 583
 FT DOMAIN 584 606
 FT DOMAIN 607 644
 FT DOMAIN 645 684
 FT DOMAIN 685 720
 FT DOMAIN 714 955
 FT DISULFID 256 256
 FT DISULFID 259 259
 FT DISULFID 285 296
 FT DISULFID 290 305
 FT DISULFID 308 319
 FT DISULFID 325 336
 FT DISULFID 330 345
 FT DISULFID 348 372
 FT DISULFID 378 392

POTENTIAL.
 THROMBOSPONDIN 4.
 N-TERMINAL.
 EGF-LIKE 1.
 EGF-LIKE 2.
 EGF-LIKE 3.
 EGF-LIKE 4.
 TSP TYPE-3 1.
 TSP TYPE-3 2.
 TSP TYPE-3 3.
 TSP TYPE-3 4.
 TSP TYPE-3 5.
 TSP TYPE-3 6.
 TSP TYPE-3 7.
 C-TERMINAL.
 INTERCHAIN (PROBABLE).
 INTERCHAIN (PROBABLE).
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.


```

RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J Shingawa A., Shibata K., Yoshino M., Itoh M., Ienii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuoka S.,
RA Atawa K., Itawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barn G.,
RA Blake J., Botfield D., Bojunga N., Carninci P., de Bonaldi M.F.,
RA Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenhack C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshewski-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohseutek S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:665-690(2001).
CC -!- FUNCTION: May be involved in RNA splicing (By similarity).
CC -!- SUBUNIT: Interacts with SRP30C, NPW1, CASP2, CASP8 and CED-3 (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 CARD domain.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK021023; BAB32281.1; -.
DR MGD; MGI:1925938; Nol3.
DR InterPro; IPR001315; CARD.
DR SMART; SMO0114; CARD; 1.
DR PROSITE; PS50209; CARD; 1.
KW Nuclear protein; mRNA splicing.
FT DOMAIN 4 95 CARD.
FT FT 132 218 GLU/PRO-RICH.
SQ SEQUENCE 220 AA; 24567 MW; A4DC057C1EB320A2 CRC64;
Query Match 20.0%; Score 117; DB 1; Length 220;
Best Local Similarity 30.7%; Pred. No. 0.0021;
Matches 31; Conservative 15; Mismatches 25; Indels 30; Gaps 5;
Oy 26 GDCCSDI-----GPMQLRELQ-----ETNALQDVYRWLRQQVREITFLKN 66
Db 121 GTTCPELP.PRASGEVGSGESSEALQRTPEPEPELEAEATGDEPD-LDGEM-----N 172
Oy 67 TWNECDACGPPOPKPOPPOPKPOPKPPPESTGSSEKD 107
Db 173 PEQEPE---PEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPD 210
RESULT 10
TONB_HELPJ STANDARD; PRT; 280 AA.
AC Q9ZJP4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TonB protein.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RA XM R.A., Ling L.-S.L., Molr D.C., King B.L., Brown E.D., Dolg P.C.,
RA Alm R.A., Ling L.-S.L., Molr D.C., deJonge B.L., Carmel G.,
RA Smith D.R., Noonan B., Guild B.C., Wria-Nickelsen M., Mills D.M., Ives C.,
RA Tummino P.J., Caruso A., Wria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson P., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001549; AAD06833.1; -.
DR PIR; F71829; F71829.
DR InterPro; IPR003538; TONB.
DR InterPro; IPR006260; TONB_C.
DR Pfam; PF03544; TonB; 1.
DR PRINTS; PR01374; TONBPROTEIN.
DR TIGRfams; TIGR01352; tonB Cterm; 1.
KW Transport; Protein transport; Inner membrane; Periplasmic;
KW Transmembrane; Signal-anchor; Repeat; Complete proteome.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 35 SIGNAL-ANCHOR (POTENTIAL).
FT DOMAIN 36 280 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 280 AA; 31045 MW; 640717A8BF81D76D CRC64;

Query Match 19.7%; Score 115; DB 1; Length 280;
Best Local Similarity 54.5%; Pred. No. 0.0039;
Matches 18; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Oy 76 POPQKPPQPOPQKPPQKPEPQSGSEKDE 108
Db 99 PKPKPKPKPKPKPKPKPKPKPKPKPKPKKE 131

RESULT 11
TONB_HELPY
D D TONB_HELPY STANDARD; PRT; 285 AA.
AC 025899;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TonB protein.
GN TONB OR HP1341.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteriaceae; Helicobacter.
OC NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilave A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,

```

[illegible]

CHROMOSOME BREAKAGE AT A SPECIFIC LOCATION; IT MAY BE THE
STRUCTURAL GENE FOR A TRANS-ACTING FUNCTION REQUIRED FOR
TRANSCRIPTION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; K01964; -; NOT_ANNOTATED_CDS.

DR PIR; A03540; TOZMCA.

DR MaltzDB; 65747; -.

DR InterPro; IPR003656; BED_finger.

DR Pfam; PF02892; 2f-BED; 1.

DR SMART; SM00614; Znf_BED; 1.

KW Transposable element; Transposition; DNA-binding; DNA recombination.

SEQUENCE 839 AA; 97068 MW; 803606AA325EA042 CRC64;

Query Match 19.3%; Score 112.5; DB 1; Length 839;
Best Local Similarity 40.4%; Pred. No. 0.019;
Matches 23; Conservative 8; Mismatches 13; Indels 13; Gaps 1;

OY 56 QOQVEIFLKNVVECDACGP-----QOPKPOPOPOPOKPKPEPE 99
ID PARX TRYB STANDARD; PRT; 115 AA.
AC P14073;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE Procylic form specific polypeptide precursor (Procyclin) (PAP).
GN PROA.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=227 (ILTart1);
RX MEDLINE=90067841; PubMed=2573878;
RA Koenig E., Delius H., Carrington M., Williams R.O., Roditi I.;
RT "Duplication and transcription of procyclin genes in Trypanosoma
RT brucei.";
RL Nucleic Acids Res. 17:8727-8739(1989).
CC -1- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
CC DIFFERENTIATION IN THE INSECT VECTOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X16015; CA34147.1; -.

DR PIR; S14896; PAUTPC.

KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.

FT CHAIN 1 27 PROCYCLIC FORM SPECIFIC POLYPEPTIDE.
FT PROPEP 28 93
FT DOMAIN 94 115
FT SIGNAL 95 92
FT CARBOHYD 59 92 17 X 2 AA TANDEM REPEATS OF [DE]-P.
FT CARBOHYD 56 56 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 115 AA; 11714 MW; 6D39F0731CF0509E CRC64;

Query Match 19.0%; Score 111; DB 1; Length 115;
Best Local Similarity 58.3%; Pred. No. 0.0035;
Matches 14; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY 76 PPOPKPOPOPOKPKPEPE 99
ID PARX TRYB STANDARD; PRT; 129 AA.
AC P09791;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Procylic form specific polypeptide A-beta precursor (Procyclin) (PAP
DE A-beta).
GN PARPA-BETA.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=87115776; PubMed=3808022;
RA Roditi I., Carrington M., Turner M.;
RT "Expression of a polypeptide containing a dipeptide repeat is
RT confined to the insect stage of Trypanosoma brucei.";
RL Nature 325:272-274(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258895; PubMed=2342468;
RA Clayton C.E., Frieri J.P., Itzhaki J.E., Bellofatto V., Sherman D.R.,
RA Wisdom G.S., Vijayasathiy S., Mowatt M.R.;
RT "Transcription of the procylic acidic repetitive protein genes of
RT Trypanosoma brucei.";
RL Mol. Cell. Biol. 10:3036-3047(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX Vijayasathiy S., Ernest I., Itzhaki J., Sherman D., Mowatt M.R.,
RA Michels P.A.M., Clayton C.E.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 28-64, AND POST-TRANSLATIONAL MODIFICATIONS.
RC STRAIN=427;
RX MEDLINE=89359323; PubMed=2475493;
RA Clayton C.E., Mowatt M.R.;
RT "The procylic acidic repetitive proteins of Trypanosoma brucei.
RT purification and post-translational modification.";
RL J. Biol. Chem. 264:15088-15093(1989).
CC -1- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
CC DIFFERENTIATION IN THE INSECT VECTOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X04814; CA28503.1; -.

DR EMBL; M33129; AAA30225.1; -.

DR EMBL; X52584; CA36815.1; -.

DR PIR; A26036; A26036.

KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.

FT CHAIN 1 27
FT SIGNAL 28 107
FT SEQUENCE 28 107 PROCYCLIC FORM SPECIFIC POLYPEPTIDE A-
FT BETA.

GenCore version 5.1.6
(c) 1993 - 2003 Compu

(without alignments)
239.016 Million cell updates/sec

US-09-696-872-23

584
1 MRYMILGLLALAVCSAKK.....PKPQKPEECTGSSEKDEL 109

BLOSUM62

830525 seqs, 258052604 residues

hits satisfying chosen parameters: 830525

length: 2000000000

Minimum Match	0%
Maximum Match	100%

Listing first 45 summaries

SPTREMBL_23:*

```

1:  SPREMBL.23.*
2:  sp archaea.*
3:  sp bacteria.*
4:  sp fungi.*
5:  sp human.*
6:  sp invertelbrate.*
7:  sp mammal.*
8:  sp_mhc.*
9:  sp_oranelle.*
10: sp_phage.*
11: sp_plant.*
12: sp_rodent.*
13: sp_virus.*
14: sp_vetelpare.*
15: sp_unclassified.*
16: sp_virus.*
17: sp_bacteriap.*
18: sp_archaeap.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	256.5	43.9	130	4	Q8N2R4	Q8N214	homo sapien
2	247.5	42.4	755	6	Q9SG80	Q9SG80	equus caball
3	245.5	42.0	817	4	O14592	O14592	homo sapien
4	236	40.4	755	11	Q9R0G6	Q9R0G6	mus musculus
5	236	40.4	755	11	Q8V154	Q8V154	mus musculus
6	195.5	33.5	724	4	Q8N4T2	Q8N4T2	homo sapien
7	153	26.2	1119	4	O94885	O94885	homo sapien
8	146.5	25.1	863	11	Q9QY53	Q9QY53	mus musculus
9	146.5	25.1	963	11	Q9Z1T2	Q9Z1T2	mus musculus
10	133	22.8	372	5	Q25649	Q25649	plasmidium
11	133	22.8	332	5	Q25649	Q25649	plasmidium
12	132	22.6	314	12	Q69088	Q69088	human herpes
13	127	21.7	2387	16	Q8PF8	Q8PF8	escherichia
14	126	21.6	416	5	Q03752	Q03752	plasmidium
15	124	21.2	910	11	Q9JL8	Q9JL8	mus musculus
16	124	21.2	912	11	Q9JL7	Q9JL7	mus musculus

Query Match	Best Local Similarity	43.9%	Score 256.5;	DB 4;	Length 130;
Matches 61;	Conservative	8;	Mismatches 23;	Indels 17;	Gaps 3;
QY	8	LLAAAVCSAKKSSIGDCCSDLGPMRELQETNALQDVBDMLQOVRREITFLKNT	67		
DB	10	LLTTLAAGASGQSGSPG---SLGPMRELQETNALQDVBDMLQOVRREITFLKNT	65		
OY	68	VMECDACGPGP-----PQPKPQPOP-----QPKPKPEEGTGS	103		
DB	66	VMECDACGMOOSVRTGTPSVRPLHCAAGVYLRGLPAGVQRPHPRGRA	114		

RESULT 2					
ID	Q9BG80	PRELIMINARY;	PRT;	755 AA.	
AC	Q9BG80;				
DT	01-JUN-2001	(TREMBLrel. 17, Created)			
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)			
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)			
DE	Cartilage oligomeric matrix protein.				
GN	COMP.				
OS	Equus caballus (Horse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Perissodactyla; Equidae; Equus.				
OX	NCBI_Taxid=9796;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Dudhia J., Williams D., Smith R.K.;				
RT	"Molecular characterization and tissue distribution of equine				
RT	cartilage oligomeric protein (Thrombospondin 5).";				
RL	Submitted (DBC-2800) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF325902; AAG59881.1; ..				
DR	HSSP; P35444; 1VDF.				
DR	InterPro; IPR002048; EF-hand.				
DR	InterPro; IPR001881; EGF Ca.				
DR	InterPro; IPR006209; EGF-like.				
DR	InterPro; IPR003367; ttp_3.				
DR	Pfam; PF02412; esp_3; 10.				
DR	SMART; SM00179; EGF_CA; 2.				
DR	PROSITE; PS00018; EF_HAND; 1.				
DR	PROSITE; PS01186; EGF_2; 1.				
DR	PROSITE; PS01187; EGF_CA; 2.				
KW	EGF-like domain; Matrix protein.				
SQ	SEQUENCE 755 AA; 82162 MW; 17CDFA4B1A79CEC7 CRC64;				
Query Match 42.4%; Score 247.5; DB 6; Length 755;					
Best Local Similarity 63.2%; Pred. No. 8,7e-17;					
Matches 55; Conservative 11; Mismatches 16; Indels 5; Gaps 2.					
QY	8 LLAIAVCSAAKKGSSLGDCSCSDIGPOMLRLOETNALODVRDWLRFQVREITFLKNT 67				
DB	10 LLALAAVLSSSGO--TPIG---TELGPOMLRLOETNALODVRLRHQQVAEIFLKNT 64				
QY	68 VMECDACGPOPKRQPOPKRQPOPKRQ 94				
DB	65 VMECDACGMPARTPRVSVRPLAQCAP 91				
RESULT 3					
ID	O14592	PRELIMINARY;	PRT;	817 AA.	
AC	O14592;				
DT	01-JAN-1998	(TREMBLrel. 05, Created)			
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)			
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)			
DE	COMP_HUMAN.				
GN	COMP.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Lamerlin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,				
RA	Gordon L., Kyle A., Ramirez M., Stillwagen S., Garnes J., Danganan L.,				
RA	Christensen M., Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A.,				
RA	Olson A.O., Carraro A.V.;				
RT	"Sequence analysis of an ~1 Mb region containing the MEZF2 gene in				
RT	19p12.";				
RL	Submitted (NCV-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AC003107; AAB86501.1; -.				
DR	HSSP; P35444; 1VDF.				
DR	InterPro; IPR002048; EF-hand.				
DR	InterPro; IPR001881; EGF Ca.				
DR	InterPro; IPR006209; EGF-like.				

```
DR InterPro; IPR003367; top_3.  
DR Pfam; PF02412; tsp_3; 11.  
DR SMART; SM00179; EGF_CA; 2.  
DR PROSITE; PS00018; EF_HAND; 1.  
DR PROSITE; PS01186; EGF_2; 1.  
DR PROSITE; PS01187; EGF_CA; 2.  
KW EGF-like domain.  
SQ SEQUENCE 817 AA; 89148 MW; 70551980BE221855 CRC64;  
  
Query Match 42.0%; Score 245.5; DB 4; Length 817;  
Best Local Similarity 67.1%; Pred. No. 1.5e-16;  
Matches 55; Conservative 6; Mismatches 10; Indels 11; Gaps 2;  
  
QY 8 LLAALAVCSAAKKSSISGDCSDLGPMIRELOETNAALODVVDMLRQGVREITFLKNT 67  
| | | : : : | | | | | | | | | | | | | | | | | | | | | | |  
DB 10 LLTFAAGASGGQSGSPIG---SDLPGMIRELOETNAALODVRELLRQGVREITFLKNT 65  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 68 VMECDACGPQ-----PQPKP 82  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 66 VMECDACGMQGSVRTGLPSVP 87  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
RESULT 4  
Q9ROG6 PRELIMINARY; PRT; 755 AA.  
ID Q9ROG6  
AC Q9ROG6;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DI 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
DE Cartilage oligomeric matrix protein precursor.  
GN COMP.  
OS Mus musculus (Mouse).  
OC Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
CX NCBI_TaxID=10090;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cartilage;  
RA Pang C., Carlson C.S., Leslie M.P., Tulli H., Stoleran E., Perris R.,  
RA Ni L., Di Cesare P.B.;  
RT "Molecular Cloning, Sequencing, Tissue and Developmental Expression of  
RT Mouse Cartilage Oligomeric Matrix Protein (COMP)".  
RL J. Orthop. Res. 0:0-0(1999).  
DR EMBL; AF033530; AAD01972.1; -.  
DR HSSP; P35444; 1VDF.  
DR MGI; MGI:88469; Comp.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001881; EGF_Ca.  
DR InterPro; IPR006209; EGF_like.  
DR InterPro; IPR003367; top_3.  
DR Pfam; PF02412; tsp_3; 9.  
DR SMART; SM00179; EGF_CA; 2.  
DR PROSITE; PS00018; EF_HAND; 1.  
DR PROSITE; PS01186; EGF_2; 1.  
DR PROSITE; PS01187; EGF_CA; 2.  
KW EGF-like domain; Matrix protein; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 755 CARTILAGE OLIGOMERIC MATRIX PROTEIN.  
SQ SEQUENCE 755 AA; 82352 MW; 7DDFCF443589A0B7 CRC64;  
  
Query Match 40.4%; Score 236; DB 11; Length 755;  
Best Local Similarity 59.8%; Pred. No. 1.3e-15;  
Matches 52; Conservative 8; Mismatches 23; Indels 4; Gaps 1;
```


ID	Q8V154	PRELIMINARY:	PRT:	755 AA.
AC	Q8V154			
DT	01-MAR-2002 (TREMBlrel. 20, Created)			
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)			
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)			
DE	Cartilage oligomeric matrix protein.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Fang C., Leslie M.P., Nord R., Tian H., Di Cesare P.E.;			
RL	"Mouse Cartilage Oligomeric Matrix Protein Genomic Sequence.";			
RT	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF257516; AAL36518.1; -			
DR	InterPro; IPR002048; EF-hand.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR003679; EGF_like.			
DR	Pfam; PF02412; tsp_3; 9..			
DR	SMART; SM00179; EGF_CA; 2.			
DR	PROSITE; PS00018; EF_HAND; 1.			
DR	PROSITE; PS01186; EGF_2; 1.			
DR	PROSITE; PS01187; EGF_CA; 2.			
KW	EGF-like domain; Matrix protein.			
SO	SEQUENCE 755 AA; 82286 MW; 91758FCF789167EB CRC64;			
Query Match	40.4%; Score 236; DB 11; Length 755;			
Best Local Similarity	59.8%; Fred. No. 1.3e-15;			
Matches	52; Conservative 8; Mismatches 23; Indels 4; Gaps 1;			
QY	8 ILAIAVSAKKSGSLGDCSDLDGPQMLKEIQTNALQDVDPWLROVREITFLKNT 67			
DB	9 VLAIALHATGCGQIPLEG---DLAPQMLRELQTNALQDVRELHVEKTEITFLKNT 64			
QY	68 VMECDACGPPQPPKPPQPPQPPKPPQPP 94			
DB	65 VMECDACGMPARTPTGTLVSRVPLCAP 91			
RESULT 6				
Q8N4T2	PRELIMINARY;	PRT;	724 AA.	
AC	Q8N4T2			
DT	01-OCT-2002 (TREMBlrel. 22, Created)			
DT	01-OCT-2002 (TREMBlrel. 22, Last sequence update)			
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)			
DE	Similar to cartilage oligomeric matrix protein (pseudochondroplasia, epiphyseal dysplasia 1, multiple).			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain, and Lung;			
RA	Straubeberg R., and Lung;			
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC033676; AAH33676.1; -			
DR	InterPro; IPR002048; EF-hand.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR006209; EGF_like.			
DR	InterPro; IPR006210; IEGF.			
DR	SMART; SM00181; EGF; 4.			
DR	SMART; SM00179; EGF_CA; 3.			
DR	PROSITE; PS00018; EF_HAND; 1.			
DR	PROSITE; PS01186; EGF_2; 1.			
DR	PROSITE; PS01187; EGF_CA; 2.			
KW	Matrix protein.			
SO	SEQUENCE 724 AA; 79696 MW; 9AE2DB3F88815FA1 CRC64;			

```

Query Match      33.5% ; Score 195.5; DB 4; Length 724;
Best Local Similarity 75.9% ; Pred. No. 1.6e-11;
Matches 41; Conservative 2; Mismatches 4; Indels 7; Gaps 1;

Qy          36 MLRELSTNNAADVDRDMLRQGVREITFLKNTWMECDACGPQ-----PQPKP 82
            |||||::|||::|||||::|||::|||::|||::|||::|||::|||::|||::|||
Db          1 MLRLQETNNALQDVRELLRQGVREITFLKNTWMECDACGMSQSVRTGLPSVRP 54

RESULT 7
O94885      PRELIMINARY; PRT; 1319 AA.
ID ID094885
AC O94885;
DT 01-MAY-1999 (TrEMBLrel.10, Created)
DT 01-MAY-1999 (TrEMBLrel.10, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel.23, Last annotation update)
DN Hypothetical protein KIAA0790 (Fragment).
GN KIAA0790.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Koehani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998)
DR EMBL, AB018333; BAA34510.1; -.
DR InterPro, IPR001660; SAM.
DR InterPro, IPR001452; SH3.
DR Pfam, PF00536; SAM; 2.
DR Pfam, PF00018; SH3; 1.
DR SMART, SMO0454; SAM; 2.
DR SMART, SMO0326; SH3; 1.
DR PROSITE, PS50105; SAM_DOMAIN; 2.
KW Hypothetical protein.
FT NON TER
FT FT
SQ SEQUENCE 1319 AA; 143613 MW; FB97E509FEDBBEF93 CRC64;

Query Match      26.2% ; Score 153; DB 4; Length 1319;
Best Local Similarity 35.5% ; Pred. No. 6.1e-07;
Matches 38; Conservative 16; Mismatches 27; Indels 26; Gaps 5;

Qy          7 GLTALAIVCSA-AKGSSSLGDCDS-----DLGPQMRELQETNALQDVDRDMLRQGVRE 60
            |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          23 GVAGAAAGCASHGARLGDAAGDPASGGAAAGCGCARARGIGRTARA----- 68

Qy          61 ITFLKNTWME-CDACGPQPQPKPQPOPQPKPQPKPEPE-GTGSSSE 105
            ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db          69 ----RDIAMEDAGAAGCPPEPEPEPEPEPEPEPEPEPEPEPEPGATGTS 111

RESULT 8
O90Y53      PRELIMINARY; PRT; 863 AA.
ID ID090Y53
AC O90Y53;
DT 01-MAY-2000 (TrEMBLrel.13, Created)
DT 01-MAY-2000 (TrEMBLrel.13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel.23, Last annotation update)
DN Thrombospondin 4 (Fragment).
GN THBS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF152393; AAD32714.1; -.
DR EMBL; AF152392; AAD32714.1; JOINED.
DR HSSP; P35444; 1VDF.
DR MGD; MGI:110179; Tnbs4.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR003167; tsp_3.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF02412; tsp_3; 9.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW EGF-like domain.
FT NON TER 1
SQ SEQUENCE 863 AA; 95339 MW; 68E3EB6846728E4D CRC64;

Query Match 25.1%; Score 146.5; DB 11; Length 863;
Best Local Similarity 37.0%; Pred. No. 1.8e-06;
Matches 30; Conservative 12; Mismatches 36; Indels 3; Gaps 1;

QY 19 KKGSSLGDCDDLPQMLRELQETNALQDVRLRQGVREITFLKNTVMECDACGPOP 78
DB 108 QQSEPLAATSTGDFNRQFLGQMTQLNLGVEVDLRLQGVKETSFLRNTIAEQACGPLS 167
QY 79 QPKPQPQ---PQPKPQPQPKP 96
DB 168 FOSPTPNTLVPIAPPAPPTRP 188

RESULT 9
Q921T2 PRELIMINARY; PRT; 963 AA.
AC Q921T2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Thombospondin-4.
GN THBS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF102887; AAC73003.1; -.
DR HSSP; P35444; 1VDF.
DR MGD; MGI:110179; Tnbs4.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR003167; tsp_3.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF02412; tsp_3; 9.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00018; EF_HAND; 1.
RX
```

```
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW EGF-like domain.
SQ SEQUENCE 963 AA; 106366 MW; B8BA83884F489FB1 CRC64;

Query Match 25.1%; Score 146.5; DB 11; Length 963;
Best Local Similarity 37.0%; Pred. No. 2e-06;
Matches 30; Conservative 12; Mismatches 36; Indels 3; Gaps 1;

QY 19 KKGSSLGDCDDLPQMLRELQETNALQDVRLRQGVREITFLKNTVMECDACGPOP 78
DB 208 QQSEPLAATSTGDFNRQFLGQMTQLNLGVEVDLRLQGVKETSFLRNTIAEQACGPLS 267
QY 79 QPKPQPQ---PQPKPQPQPKP 96
DB 268 FOSPTPNTLVPIAPPAPPTRP 288

RESULT 10
Q25648 PRELIMINARY; PRT; 272 AA.
AC Q25648;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Circumsporozoit (CS) protein (Fragment).
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87216962; PubMed=3556207;
RA Weber J.L., Egan J.E., Lyon J.A., Wirtz R.A., Charoenvit Y.,
RA Maloy W.L., Hockmeyer W.T.;
RT "Plasmodium berghei: Cloning of the circumsporozoit protein gene.";
RL Exp. Parasitol. 63:295-300(1987).
DR EMBL; M25445; AAA29531.1; -.
DR InterPro; IPR003067; Circmeprozite.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR PRINTS; PR01303; CRCSPPRZOTE.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER 1
SQ SEQUENCE 272 AA; 29408 MW; 4FF07FA62B32A051 CRC64;

Query Match 22.8%; Score 133; DB 5; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.2e-05;
Matches 20; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 76 PQPKPQPQPKPQPKPQPKPEPGTGSSEKD 107
DB 149 PQPKPQPQPKPQPKPQPKPGGNNNNKVN 180

RESULT 11
Q25649 PRELIMINARY; PRT; 332 AA.
AC Q25649;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Circumsporozoit protein.
GN CSP.
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NK65;
RX MEDLINE=90158693; PubMed=2406593;
RX
```

RA Lanar D.E.;
 RT "Sequence of the circumsporozoite gene of Plasmodium berghel ANKA
 RL clone and NK65 strain.";
 RL M01. Biochem. Parasitol. 39:151-154(1990).
 DR EMBL; M28887; AAA29541.1; -.
 DR InterPro; IPR003067; Circmsprzoite.
 DR InterPro; IPR002965; P_rich_extenun.
 DR InterPro; IPR000884; TSP1.
 DR Pfam; PF00090; tsp_1; 1.
 DR PRINTS; PR01303; CRCMSPRZOTE.
 DR PRINTS; PR01217; FRICHEXTENSN.
 DR SMART; SMO0209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 SQ SEQUENCE 332 AA; 36196 MW; 07C2D2F265EAAEB CRC64;

Query Match 22.8%; Score 133; DB 5; Length 332;
 Best Local Similarity 62.5%; Pred. No. 1.5e-05;
 Matches 20; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Oy 76 PQPQPKPQPPQPKPQPKPEEGTGSSEKD 107
 Db 209 PQPQPKPQPPQPKPQPKPEEGTGSSEKD 240

RESULT 12

ID 069088 PRELIMINARY; PRT; 3164 AA.
 AC 069088;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Viton protein.
 OS Human herpesvirus 1.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxID=10298;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17;
 RX MEDLINE=88274327; PubMed=2839594;
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNaab D., Perry L.J., Scott J.E., Taylor P.;
 RT "The complete DNA sequence of the long unique region in the genome of
 RT herpes simplex virus type 1.";
 RL J. Gen. Virol. 69:1531-1574(1988).
 DR EMBL; D10879; BA01682.1; -.
 DR InterPro; IPR006928; Herpes_teg_N.
 DR InterPro; IPR005210; Herpes_UL36.
 DR Pfam; PF04843; Herpes_teg_N; 1.
 DR Pfam; PF03586; Herpes_UL36; 1.
 SQ SEQUENCE 3164 AA; 335817 MW; 2E3335F9525F8C71 CRC64;

Query Match 22.6%; Score 132; DB 12; Length 3164;
 Best Local Similarity 66.7%; Pred. No. 0.00021;
 Matches 22; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 67 TMECDACGPPQPKPQPKPQPKPEPE 99
 Db 2904 TVSRSLAPQPPQPKPQPKPQPKPEPE 2936

RESULT 13

ID 08FF8 PRELIMINARY; PRT; 2387 AA.
 AC 08FF8;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Yaph homolog.
 GN C2895.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=06:HI / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackert J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL; AB016764; AAN81345.1; -.
 KW Complete proteome.
 SQ SEQUENCE 2387 AA; 242283 MW; 06298DC5293DD03 CRC64;

Query Match 21.7%; Score 127; DB 15; Length 2387;
 Best Local Similarity 76.9%; Pred. No. 0.0005;
 Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 76 PQPQPKPQPPQPKPQPKPEEGT 101
 Db 2291 PQPQPKPQPPQPKPQPKPEEGT 2316

RESULT 14

ID 003752 PRELIMINARY; PRT; 416 AA.
 AC 003752;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Circumsporozoite protein.
 GN CS.
 OS Plasmodium yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5861;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9114845; PubMed=2290446;
 RA Colomer-Gould V., Enea V.;
 RT "Plasmodium yoelii nigeriensis circumsporozoite gene structure and its
 RT implications for the evolution of the repeat regions.";
 RL M01. Biochem. Parasitol. 43:51-58(1990).
 DR EMBL; M58295; AAA29528.1; -.
 DR InterPro; IPR003067; Circmsprzoite.
 DR InterPro; IPR000884; TSP1.
 DR Pfam; PF00090; tsp_1; 1.
 DR PRINTS; PR01303; CRCMSPRZOTE.
 DR SMART; SMO0209; TSP1; 1.
 DR PROSITE; PS50092; TSP1; 1.
 SQ SEQUENCE 416 AA; 45135 MW; BD8CGAD168BEB7F0 CRC64;

Query Match 21.6%; Score 126; DB 5; Length 416;
 Best Local Similarity 79.2%; Pred. No. 9.8e-05;
 Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Oy 76 PQPQPKPQPPQPKPQPKPEPE 99
 Db 280 PQPQPKPQPPQPKPQPKPEPE 303

RESULT 15

ID 09JL8 PRELIMINARY; PRT; 910 AA.
 AC 09JL8;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE GABA-A receptor epsilon-like subunit.
 GN GABRE OR EPSILON.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20266167, Pubmed=10804200;
RA Sankkonen S.T., Hanna M.C., Kirkness E.F., Korpi E.R.;
RT "GABA receptor epsilon and theta subunits display unusual structural
RT variation between species and are enriched in the rat locus
 cereulus." ;
RL J. Neurosci. 20:3588-3595(2000).
CC -1. SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1. SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
DR EMBL; AF189263; GAEF0384.1; -.
DR MGD; MGI:1330235; Aabre.1.
DR InterPro; IPR006201; Neur channel.
DR InterPro; IPR006202; Neur channel LBD.
DR InterPro; IPR006029; Neur channel memb.
DR InterPro; IPR005162; PpanTne_atetch.
DR Pfam; PF02931; Neur_chan_LBD; 1.
DR Pfam; PF02932; Neur_chan_memb; 1.
DR PRINTS; PR00252; NRIONCHANNEL.
DR PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
DR PROSITE; PS00012; PHOSPHORANETHYLIN; 1.
KW Glycoprotein; Ionic channel; Postsynaptic membrane; Receptor;
KW Transmembrane.
SQ SEQUENCE 910 AA; 103801 MW; 65052571EC8CECBF CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 15:53:48 ; Search time 4.28319 Seconds
(without alignments)
148.232 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 KDELL 4

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	4	AA41971	Mammalian endoplas
2	20	100.0	4	AA48246	Endoplasmic reticu
3	20	100.0	4	AA49442	Immunomodulatory c
4	20	100.0	4	AA49584	Sequence of endopl
5	20	100.0	4	AA95219	Pseudomonas exotox
6	20	100.0	4	AA94944	Microsomal retenti
7	20	100.0	4	AA95063	Mammalian endoplas
8	20	100.0	4	AAW31861	Endoplasmic reticu
9	20	100.0	4	AAW21969	Linker #3 for immu

10	20	100.0	4	AAW76396	Endoplasmic reticu
11	20	100.0	4	AAW64614	Endomembrane reten
12	20	100.0	4	AAW61591	Endomembrane reten
13	20	100.0	4	AAW65513	Endoplasmic reticu
14	20	100.0	4	AAW51434	C-terminal sequenc
15	20	100.0	4	AAW51437	C-terminal sequenc
16	20	100.0	4	AAW52264	C-terminal fragmen
17	20	100.0	4	AB876201	Peptide used to mo
18	20	100.0	4	AAV49700	Pseudomonas exotox
19	20	100.0	4	AAV29637	KDEL peptide for t
20	20	100.0	4	AAV30052	Organelle targetin
21	20	100.0	4	AAV25688	Membrane retaining
22	20	100.0	4	AAV27177	Organelle-targetin
23	20	100.0	4	AAW85767	Endomembrane reten
24	20	100.0	4	AAW86345	Endoplasmic reticu
25	20	100.0	4	AAW82925	Pseudomonas sp. ex
26	20	100.0	4	AAW92911	Pseudomonas sp. ex
27	20	100.0	4	AAW92046	CD4+ T-cell activa
28	20	100.0	4	AAW30285	Targeting and inte
29	20	100.0	4	AAW09912	Endoplasmic reticu
30	20	100.0	4	AAW03195	Amino acid sequenc
31	20	100.0	4	AAV93738	Protein localisati
32	20	100.0	4	AAV98152	Endoplasmic reticu
33	20	100.0	4	AAV92046	Endoplasmic reticu
34	20	100.0	4	AAV70697	Rat endoplasmic re
35	20	100.0	4	AAV90442	Mammalian cytoplas
36	20	100.0	4	AAV69694	Mammalian cytoplas
37	20	100.0	4	AAV68883	Endoplasmic reticu
38	20	100.0	4	AAV32395	Endoplasmic reticu
39	20	100.0	4	ABG99144	Calreticulin endop
40	20	100.0	4	AAU07199	Human ICAW-1 extra
41	20	100.0	4	AAW67291	Amino acid sequenc
42	20	100.0	4	AAW678315	ER lumen retention
43	20	100.0	4	AAW05817	Protein localisati
44	20	100.0	4	AAU03564	Peptide retention
45	20	100.0	4	AAW43355	Amino acid sequenc
			4	AAW89995	C-terminal ER rete

ALIGNMENTS

RESULT 1
AA41971
ID: AA41971 standard; peptide; 4 AA.
XX
AC AA41971;
XX
DT 25-MAR-2003 (updated)
DT 10-MAR-2003 (updated)
DT 22-OCT-1993 (first entry)
XX
DE Mammalian endoplasmic reticulum retention signal.
XX
KW yeast expression cassette; yeast alpha-factor leader; KEX2p;
KW heterologous protein production; mating pheromone; YAP3;
XX
KW diabolic processing endoprotease.
XX
OS Mammalia.
XX
FN EP548012-A1.
XX
PD 23-JUN-1993.
XX
PF 08-DEC-1992; 92EP-0810964.
XX
PR 16-DEC-1991; 91EP-0810984.
XX
PA (CIBA) CIBA GEIGY AG.
PA (NOVS) NOVARTIS AG.
XX
PI Chaudhuri B, Rietman H, Seeboth P, Stephan C;
XX WPI; 1993-198976/25.
DR

XX Endoplasmic reticulum-located recombinant di:baeic endoprotease -
 PT used in host cells to raise ratio of active, correctly folded
 PR heterologous protein
 XX
 XX
 PS Disclosure; Page 3; 30pp; English.
 CC
 CC Dibaic endoproteases of the invention are localised to the
 CC endoplasmic reticulum of yeast host cells via ER-retention signals.
 CC A preferred retention signal is the tetrapeptide KDEL functional in
 CC mammalian cells. See also AA042671, AA042672 and AA041971-2.
 CC (Updated on 10-MAR-2003 to add missing OS field.)
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 CC
 SQ Sequence 4 AA;
 Query Match 100.0%; Score 20; DB 14; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KDEL 4
 DB 1 KDEL 4
 RESULT 2
 ID AA048246 standard; peptide; 4 AA.
 AC
 AC AA048246;
 XX
 XX 25-MAR-2003 (updated)
 DT 29-JUL-1994 (first entry)
 XX
 XX Endoplasmic reticulum retention signal KDEL.
 DE
 XX
 XX Single chain antibody; sFv; heavy chain; light chain; kappa;
 KW variable domain; hydrophilic linker; antibodies;
 KW endoplasmic reticulum retention.
 XX
 OS Synthetic.
 OS
 PI WO9402610-A1.
 PN
 XX 03-FEB-1994.
 XX
 XX 16-JUL-1993; 93WO-US06735.
 PF
 XX 17-JUL-1992; 92US-0916939.
 PR 17-MAR-1993; 93US-0045274.
 XX
 XX (DAND) DANA FARMER CANCER INST INC.
 PA
 PI Haeeltine WA, Marasco WA;
 XX
 XX WPI; 1994-048868/06.
 DR
 XX Intracellular binding of antigens - by using antibody targeting
 PT with vector system, for e.g. tumour suppression
 PS
 PS Claim 3; Page 99; 155pp; English.
 CC
 CC New vector systems comprise a sequence adapted for intracellular
 CC delivery and expression contg. a promoter operably linked to an
 CC antibody gene encoding an antibody which binds to a specific target
 CC antigen. The antibody is esp. a single chain antibody in which the
 CC heavy and light chain variable regions are joined via a hydrophilic
 CC linker peptide. Localisation sequences are pref. included in the
 CC constructs. See AA048246-9 and AA048252-3 for pref. (known) endoplasmic
 CC reticulum retention signals.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC

SQ Sequence 4 AA;
 Query Match 100.0%; Score 20; DB 15; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KDEL 4
 DB 1 KDEL 4
 RESULT 3
 ID AA049442 standard; Protein; 4 AA.
 AC
 AC AA049442;
 XX
 XX 25-MAR-2003 (updated)
 DT 16-SEP-1994 (first entry)
 XX
 XX Immunomodulatory trafficking sequence #1.
 DE
 XX
 XX Naturally-occurring; immunomodulatory protein; human; therapy; class I;
 KW major histocompatibility complex; class II; allotype; type I diabetes;
 KW autoimmune disease; rheumatoid arthritis; T-cell-mediated response;
 KW multiple sclerosis; transplant rejection; vaccine; MHC.
 XX
 OS Homo sapiens.
 OS
 PI WO9404171-A1.
 PN
 XX 03-MAR-1994.
 XX
 XX 11-AUG-1993; 93WO-US07545.
 PF
 XX 11-AUG-1992; 92US-0925460.
 PR 15-JUN-1993; 93US-0077253.
 XX
 XX (HARD) HARVARD COLLEGE.
 PA
 PI Chicz RM, Hedley ML, Stern LJ, Strominger JL, Urban RG;
 PI Vignali DA;
 XX
 XX WPI; 1994-082825/10.
 DR
 XX
 XX Novel immunomodulatory peptide(s) and nucleic acids - useful for
 PT treatment of auto-immune diseases, transplant rejection and for
 PT vaccination
 PS
 PS Claim 13; Page 94; 139pp; English.
 CC
 CC The sequences given in AA049291-505 and AA046981-7038 represent peptide
 CC fragments of naturally-occurring immunomodulatory proteins. These
 CC fragments are between 10-30 residues in length and bind to a human
 CC major histocompatibility complex (MHC) class II allotype. These
 CC peptides may be used for therapy of autoimmune diseases, such as
 CC type I diabetes, rheumatoid arthritis and multiple sclerosis, and to
 CC reduce transplant rejection. They may also be used for vaccination
 CC providing an exclusively T-cell-mediated response, which can be
 CC class I or class-II based, or both, depending on the length and
 CC character of the immunogenic peptides.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC (Updated on 25-MAR-2003 to correct PR field.)
 CC
 SQ Sequence 4 AA;
 Query Match 100.0%; Score 20; DB 15; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KDEL 4
 DB 1 KDEL 4

RESULT 4
 ID AAR49584 standard; peptide; 4 AA.
 XX AAR49584;
 AC AAR49584;
 DT 25-MAR-2003 (updated)
 DT 15-SEP-1994 (first entry)
 XX Sequence of endoplasmic reticulum (ER) retention peptide.
 DE Endoplasmic reticulum; ER; trafficking sequence.
 XX Synthetic.
 OS WO9404557-A1.
 PN 03-MAR-1994.
 PD 11-AUG-1992; 92WO-US06692.
 PF 11-AUG-1992; 92WO-US06692.
 PR 11-AUG-1992; 92WO-US06692.
 XX (HARD) HARVARD COLLEGE.
 PA Chicz, RM, Hedley ML, Stern LJ, Strominger JL, Urban RG;
 PI Vignali DM;
 XX WPI; 1994-083102/10.
 DR WPI; 1994-083102/10.
 PT New peptide binding to MHC class II allotype - useful for
 PT treating auto-immune diseases, transplant rejection and for
 PT immunisation
 XX Claim 20; Page 49; 60pp; English.
 PS A trafficking sequence is an AA sequence which functions to control
 CC intracellular trafficking (directed movement from organelle to
 CC organelle or to the cell surface) of a polypeptide to which it is
 CC attached. Such trafficking sequences might traffic the polypeptide
 CC to ER, a lysosome, or an endosome, and include signal peptides, ER
 CC retention peptides such as AAR49584; and lysosome-targeting peptides
 CC such as AAR4585 and AAR49586. An example of a signal peptide with such
 CC a function is the signal peptide of MHC class II alpha (AAR49587).
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX Sequence 4 AA;
 SQ

Query Match 100.0%; Score 20; DB 15; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KDEL 4
 ||||
 DB 1 KDEL 4

RESULT 5
 ID AAR95219 standard; protein; 4 AA.
 XX AAR95219;
 AC AAR95219;
 DT 16-DEC-1996 (first entry)
 DT Pseudomonas exotoxin modified C-terminal sequence.
 DE Antibody; fusion protein; single chain; inhibition; tumour;
 DE diagnosis; detection; imaging; immunotoxin; targeting; assay;
 KW immunosassay; Lewis(Y) carbohydrate antigen.
 XX

OS Pseudomonas sp.
 XX WO9613594-A1.
 PN 09-MAY-1996.
 PD 26-OCT-1995; 95WO-US13811.
 PF 28-OCT-1994; 94US-0311398.
 PR 28-OCT-1994; 94US-0311398.
 PR 28-OCT-1994; 94US-0311397.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Benhar I, Brinkmann U, Fitzgerald D, Jung S, Lee B;
 PI Padlan EA, Pai L, Pascan I, Willingham M;
 XX WPI; 1996-251462/25.
 DR Single chain fusion proteins and antibodies - useful to diagnose and
 PT treat cancer, specifically bind Lewis(Y) related carbohydrate
 PT antigen
 XX Disclosure; Page 14; 116pp; English.
 PS A novel recombinant DNA molecule which encodes a single chain fusion
 CC protein or antibody comprising the Fv region of both the light and
 CC heavy chains of an antibody (Ab) fused together, and an effector
 CC molecule, where the fusion protein or Ab has the binding specificity
 CC of monoclonal Ab (MAb) B1, B3 or B5, can be used for the production
 CC of such fusion proteins or antibodies. The fusion proteins can be
 CC used in compositions as an immunotoxin to inhibit tumour cell growth.
 CC The single chain antibody can be used to detect the presence or
 CC absence of cells bearing a Lewis(Y) carbohydrate antigen in a
 CC patient. The antibodies are also useful as multiple targeting
 CC moieties, providing at least 2 kinds of biological activity. They
 CC can also be used in diagnostic assays and for the imaging of tumours
 CC when attached to a radiolabel and for the pathological diagnosis of
 CC tumours. Humanised antibodies are less immunogenic than the mouse
 CC MAbs B1, B3 and B5, making them more suitable for long term
 CC treatment. The effector molecule used is preferably a drug or
 CC cytotoxin which then produces an immunotoxin capable of selectively
 CC killing particular cells. Preferred toxins are the Pseudomonas
 CC exotoxin or Diphtheria toxin. These are both highly toxic compounds
 CC and so are modified to eliminate domain Ia of the Pseudomonas toxin,
 CC various amino acid deletions in domains II and III, single amino
 CC acid substitutions and addition of one or more sequences at the
 CC C-terminal end. The wild type C-terminal sequence of the Pseudomonas
 CC exotoxin is given in AAR95221.
 XX Sequence 4 AA;
 SQ

Query Match 100.0%; Score 20; DB 17; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KDEL 4
 ||||
 DB 1 KDEL 4

RESULT 6
 ID AAR94944 standard; Protein; 4 AA.
 XX AAR94944;
 AC AAR94944;
 DT 31-OCT-1996 (first entry)
 DT Microsomal retention signal.
 DE Toxin; subunit; vaccine; transgenic plant; immunogen; antigen;
 KW adjuvant; immunisation.
 KW

XX OS Synthetic.
 XX PN WO9612801-A1.
 XX PD 02-MAY-1996.
 XX PF 24-OCT-1995; 95WO-US13376.
 XX PR 24-OCT-1994; 94US-0328716.
 XX PA (TULSA) TULANE EDUCATIONAL FUND.
 XX PA (TEXA) UNIV TEXAS A & M SYSTEM.
 XX PI Arntzen CJ, Clements JD, Haq TA, Mason HS;
 XX DR WPI; 1996-230602/23.
 XX PT Transgenic plants contg. E. coli heat labile enterotoxin subunits -
 XX PS used as oral vaccines for animals which consume the plant
 XX SQ Example 13; Page 95; 130pp; English.

CC A transgenic plant comprising or expressing a DNA sequence encoding
 CC an immunogenic agent can be used as an oral vaccine for animals.
 CC The vaccine is administered by the oral consumption of the plant and
 CC provides the first known functional method for immunising animals
 CC using transgenic plants, where the plants express bacterial antigens
 CC that act as both immunogens and adjuvants. The method provides an
 CC inexpensive production and delivery system for such antigens to
 CC animals. This sequence acts as a microosomal retention signal and was
 CC used to modify the LT-B Escherichia coli toxin subunit C-terminal
 CC sequence. The sequence coding for the whole LT-B gene was used in the
 CC construction of such a transgenic plant. The immunogenic agent
 CC preferably comprises the LT-B or CT-B (cholera toxin B subunit) or
 CC optionally LT-A or CT-A.

XX SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 17; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
 ||||
 DB 1 KDEL 4

RESULT 7

AA95063 AAR95063 standard; peptide; 4 AA.

AC AAR95063;

DT 22-AUG-1996 (first entry)

DE Mammalian endoplasmic reticulum retention signal.

KW Nucleic acid transfer system; gene transfer; gene therapy;
 cell targeting; multidomain protein; vector; cancer;

KW endoplasmic reticulum.

OS Synthetic.

PN WO9613599-A1.

PD 09-MAY-1996.

PF 31-OCT-1995; 95WO-EP04270.

PR 01-NOV-1994; 94EP-0810627.

PA {WELLS/} WELLS W.

XX XX Fominaya J, Wells W;
 XX PI
 XX DR WPI; 1996-239505/24.

PT Nucleic acid transfer system for gene therapy, e.g. against cancer
 PT - includes toxin translocation domain to target nucleic acid to
 PT specific cell

XX PS Disclosure; Page 11; 106pp; English.

CC Endoplasmic reticulum retention signals include a mammalian
 CC signal (AAR95063), a bacterial signal (AAT05064) from Pseudomonas
 CC exotoxin, and a yeast signal (AAR95065). The signal may form
 CC part of a multidomain protein (see also AAR95053-58) that is used
 CC with an effector nucleic acid for the transfer of nucleic acids
 CC to targeted cells as a means of gene therapy. The endoplasmic
 CC reticulum retention signal functions to affect intracellular
 CC routing of the internalized protein/nucleic acid complex.

XX SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 17; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
 ||||
 DB 1 KDEL 4

RESULT 8
 AAM31861 AAM31861 standard; peptide; 4 AA.

AC AAM31861;

DT 06-MAY-1998 (first entry)

DE Endoplasmic reticulum retention peptide.

KW RNA-loaded antigen presenting cell; trafficking sequence; APC production;
 KW tumour formation; pathogen infection; antigenic epitope; immune response;
 KW T cell proliferation; cytotoxic T lymphocyte; adoptive immunotherapy;
 KW therapy; TAB; CTL; PAB; endoplasmic reticulum retention peptide.

OS Synthetic.

PN WO9741210-A1.

PD 06-NOV-1997.

PF 30-APR-1997; 97WO-US07317.

PR 30-APR-1996; 96US-0640444.

PA (UYDU-) UNIV DUKE.

PI Boczkowski DJ, Gilboa E, Nair SK;

DR WPI; 1997-549715/50.

PT Use of RNA-loaded antigen presenting cells - to induce T-cell
 PT proliferation for the treatment or prevention of tumour formation or
 PT pathogen infection

PS Claim 49; Page 38; 47pp; English.

CC This sequence represents a endoplasmic reticulum retention peptide, and
 CC can be used in the method of the invention. The method is for producing
 CC an RNA-loaded antigen presenting cell (APC) that presents on its surface
 CC a tumour or pathogen antigenic epitope (TAB or PAB respectively) that
 CC induces T cell proliferation and an immune response against the tumour or

CC pathogen, and comprises introducing into an APC in vitro, RNA that
CC encodes the antigen. The RNA-loaded APCs can be used to stimulate
CC cytotoxic T lymphocyte (CTL) proliferation ex vivo or in vivo. The ex
CC vivo expanded CTL can be administered to a patient in a method of
CC adoptive immunotherapy. The methods can be used for treating or
CC preventing tumour formation or pathogen infection caused by e.g. HIV,
CC hepatitis, influenza, poliomyelitis, measles, herpes, mumps or rubella
CC viruses, Salmonella, Shigella or Enterobacter. The method circumvents the
CC need to purify RNA or isolate and identify a TAE or PAE.

XX Sequence 4 AA;

SO Query Match 100.0%; Score 20; DB 18; Length 4;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 9

AAW21969 standard; peptide; 4 AA.

AAW21969;

03-DEC-1997 (first entry)

Linker #3 for immunotoxin containing Pseudomonas exotoxin.

PCR; primer; amplify; polymerase chain reaction; antibody; immunotoxin;
variable heavy chain; VH; murine monoclonal antibody; Lewis; carcinoma;
carboxylate antigen; Pseudomonas exotoxin; proteolytic activation;
cytotoxic activity; tumour; autoimmune condition; rheumatoid arthritis;
graft versus host disease; organ transplant rejection; type I diabetes;
multiple sclerosis; systemic lupus erythematosus; myasthenia gravis;
T cell; B cell; cytosol; bone marrow; transplant; therapy.

OS Synthetic.

MO9713529-A1.

17-APR-1997.

11-OCT-1996; 96WO-US16327.

13-OCT-1995; 95US-0005388.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Kuan C, Pastan I;

WPI; 1997-235666/21.

Immunotoxin(s) comprising Pseudomonas exotoxin linked to
disulphide stabilised variable heavy and light chain regions of an
antibody - useful for killing target cells bearing characteristic
marker

Claim 9; Page 50; 64pp; English.

AAW21967-W21969 represent linkers used in the immunotoxins of the
invention. The immunotoxins bind to target cells, and comprise, a
Pseudomonas exotoxin (PE) that does not need proteolytic activation for
cytotoxic activity fused to a VH framework region of an Fv antibody (Ab)
fragment. The VH chain region is bound through at least one disulphide
bond to a variable light (VL) chain framework region. The PE is lacking
residues 1-279 and is at least 10-fold more cytotoxic to the target
cells than an immunotoxin comprising PE attached to a VH chain framework
region of an Fv Ab fragment lacking a disulphide bond to a VL chain
framework region. These sequences are used to join the VH chain region to
the PE. The immunotoxins can be used for killing target cells in the

CC treatment of tumours, autoimmune conditions, graft versus host disease,
CC organ transplant rejection, type I diabetes, multiple sclerosis,
CC rheumatoid arthritis, systemic lupus erythematosus, myasthenia gravis,
CC etc, all caused by T and B cells. They can also be used to deliver an
CC antibody to the cytosol of a cell, and in vitro in the elimination of
CC harmful cells from bone marrow before transplant. The immunotoxins have
CC high cytotoxicity to target cells and a small size to provide greater
CC penetration to target cells.

XX Sequence 4 AA;

SO Query Match 100.0%; Score 20; DB 18; Length 4;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 10

AAW76396 standard; Peptide; 4 AA.

AAW76396;

11-JAN-1999 (first entry)

Endoplasmic reticulum transport signal consensus.

Exotoxin A; ETA; drug delivery; membrane transport;

endoplasmic reticulum.

OS Synthetic.

MO9842876-A1.

01-OCT-1998.

24-MAR-1998; 98WO-US05710.

26-MAR-1997; 97US-0042056.

(TEXA) UNIV TEXAS SYSTEM.

Draper RK;

WPI; 1998-532023/45.

Delivering compounds to cells as new conjugate with detoxified
exotoxin A - able to cross membranes and deliver to the cytoplasm,
e.g. nucleic acids, antibodies, tumour suppressors etc.

Disclosure; Page 8; 76pp; English.

This peptide is the consensus intracellular transport signal used
for returning back to the endoplasmic reticulum (ER) proteins that
have escaped the ER and entered to Golgi complex. The mechanism
involves the KDEL receptor. The KDEL motif can replace a REDL
motif (see AAW76395) present at the C-terminal end of Pseudomonas
aeruginosa exotoxin A (ETA, see AAW76391). This motif is important
in the intracellular transport and cytotoxicity of ETA. This
suggests that ETA interacts with the KDEL receptor and that ETA
may reach the interior of the ER before penetrating the cytosol.
The invention provides a means of delivering compounds to cells as
conjugates with modified ETA. The modified ETA is able to cross
membranes and deliver e.g. therapeutic agents to the cytoplasm,
such as nucleic acids, peptides, peptide nucleic acids, single
chain antibodies and tumour suppressors.

XX Sequence 4 AA;

SO Query Match 100.0%; Score 20; DB 19; Length 4;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
 ||||
DB 1 KDEL 4

RESULT 11

AAW64614 standard; Protein; 4 AA.

AC AAW64614;

DT 03-NOV-1998 (first entry)

DE Endomembrane retention signal peptide.

KW Target; imaging agent; epithelium; transepithelial transport; diagnosis;

KM transcytosis; disease; basolateral; internalisation; J chain.

OS Unknown.

PN MO9830591-A1.

PD 16-JUL-1998.

PF 09-JAN-1998; 98WO-US00339.

PR 10-JAN-1997; 97US-0782480.

PA (EPIC-) EPICYTE PHARM INC.

PI Fitchen JH, Hein MB, Hiatt AC;

DR WPI; 1998-399066/34.

PT New epithelial tissue targeting agent - used to deliver imaging
PT agents to an epithelial surface for internalisation; useful in
PT diagnosis

PS Example 1c; Page 90; 118pp; English.

CC This sequence represents an endomembrane retention signal which is used
CC in a method involving the construction of a target molecule from human J
CC chain protein fragments. This construct is used in a method to target
CC imaging agents to epithelial surfaces at which they may remain or undergo
CC transepithelial transport via transcytosis. At least one imaging agent is
CC linked to the targeting molecule comprising a polypeptide that (a) forms
CC a closed covalent loop, (b) contains at least 3, preferably 4, peptide
CC domains having beta-sheet character separated by domains lacking
CC beta-sheet character and (c) is not full length dimeric IGA. The imaging
CC agents are useful in the diagnosis of disease. The target molecule is
CC also capable of specifically binding to a basolateral factor associated
CC with an epithelial surface to cause internalisation of a biological agent
CC linked to the target molecule.

SO Sequence 4 AA;

Query Match 100.0%; Score 20; DB 19; Length 4;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
 ||||
DB 1 KDEL 4

RESULT 12

AAW61591 standard; Peptide; 4 AA.

AC AAW61591;

XX 27-OCT-1998 (first entry)

XX Endomembrane retention signal.

XX J chain; targeting molecule; epithelial; beta-sheet; asthma;

XX cancer; inflammatory disorder; autoimmune disorder; celiac disease;

XX colitis; pneumonia; cystic fibrosis.

OS Synthetic.

PN MO9830592-A1.

PD 16-JUL-1998.

PF 09-JAN-1998; 98WO-US00542.

PR 10-JAN-1997; 97US-0782481.

PA (EPIC-) EPICYTE PHARM INC.

PI Fitchen JH, Hein MB, Hiatt AC;

DR WPI; 1998-399067/34.

PT New epithelial tissue targeting agent - used to deliver
PT biologically active compounds to an epithelial surface for
PT internalisation

PS Example 1; Page 48; 142pp; English.

CC The endomembrane retention signal is used in the synthesis of a targeting
CC molecule (TM). The TMs are used to target biological agents to
CC epithelial surfaces at which they can be internalised. The TMs comprise
CC a polypeptide that: (a) forms a closed covalent loop; (b) contains at
CC least 3, preferably 4, peptide domains having beta-sheet character
CC separated by domains lacking beta-sheet character; and (c) is not full
CC length dimeric IGA. The TMs are useful to prevent and/or treat diseases
CC associated with epithelial surfaces, e.g. asthma, cancer,
CC (myco)bacterial, viral or fungal infection, inflammatory disorders,
CC autoimmune disorders, celiac disease, colitis, pneumonia and cystic
CC fibrosis.

SO Sequence 4 AA;

Query Match 100.0%; Score 20; DB 19; Length 4;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
 ||||
DB 1 KDEL 4

RESULT 13

AAW56513 standard; Peptide; 4 AA.

AC AAW56513;

DT 28-SEP-1998 (first entry)

DE Endoplasmic reticulum recycling peptide.

KW Intracellular targeting sequence; endoplasmic reticulum;

KM DNA vaccine; genetic immunisation; allergy; autoimmune disease;

OS Synthetic.

PN MO9817323-A1.

PD 30-APR-1998.

XX 23-OCT-1997; 97WO-US19545.
 PF 23-OCT-1996; 96US-0029592.
 PR (TYPE-) UNIV PENNSYLVANIA.
 XX Madalo M, Weiner DB, Williams WV;
 XX WPI; 1998-261198/23.
 DR
 XX Plasmid encoding immunogenic target protein - used in, e.g.
 PT protective or therapeutic vaccines against allergy, cancer,
 PT microbial infection or autoimmune disease
 PS Claim 7; Page 62; 84pp; English.
 XX This peptide, when joined to the C-terminus of a protein, acts as
 CC an endoplasmic reticulum (ER) recycling signal that localises the
 CC protein from the Golgi body back to the ER when expressed in a
 CC cell. A claimed novel plasmid comprises a sequence, linked to
 CC regulatory elements, that encodes an immunogenic target protein
 CC that includes, or is linked to, an intracellular targeting sequence
 CC (ITS) such as this peptide. Other ITS peptides (see AAM56512 and
 CC AAM56514) are provided that direct localisation in the lysosome or
 CC the ER. The novel plasmid is used as a protective or
 CC therapeutic DNA vaccine to immunise against the immunogenic target
 CC protein (claimed), particularly in cases of allergy, cancer (or
 CC other proliferative diseases such as psoriasis), microbial
 CC infection or autoimmune disease, e.g. rheumatoid arthritis,
 CC insulin-dependent diabetes mellitus, Crohn's disease, asthma,
 CC penicillin anaemia and many others. A particular use is against
 CC pathogens that, at least for part of their life cycle, are
 CC intracellular, e.g. hepatitis C, HIV, Neisseria gonorrhoeae,
 CC listeria and Shigella. The cytotoxic T cell response is enhanced
 CC by delivering the immunogenic target protein to the ER.
 CC
 SQ Sequence 4 AA;
 Query Match 100.0%; Score 20; DB 19; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 KDEL 4
 DB 1 KDEL 4
 RESULT 14
 AAM51434
 ID AAM51434 standard; peptide; 4 AA.
 AC AAM51434;
 XX
 XX 02-SEP-1998 (first entry)
 DT
 XX
 DE C-terminal sequence of localisation peptides for the ER.
 XX
 KW Bifunctional chelating polysaccharide; targeting peptide; heparin;
 KW nuclear localisation; fluorescent indicator; dextran; ficol; glycogen;
 KW amylopectin; mannan; inulin; starch; agarose; cellulose; ER;
 KW endoplasmic reticulum.
 XX
 OS Synthetic.
 XX
 XX US5773227-A.
 PN
 XX 30-JUN-1998.
 PD
 XX 23-JUN-1993; 93US-0082269.
 PF
 XX 23-JUN-1993; 93US-0082269.
 PR
 XX

PA (MOLE-) MOLECULAR PROBES INC.
 XX Albritton NL, Kuhn MA, Meyer T;
 PI WPI; 1998-386995/33.
 DR
 XX Bifunctional chelating polysaccharides used for analysing
 PT intracellular ion levels - have chelating moieties attached which
 PT act as a fluorescent indicator in the presence of selected metal
 PT ions
 XX
 XX Disclosure; Column 9; 29pp; English.
 PS
 XX The invention relates to a water soluble polysaccharide with a molecular
 CC weight greater than 1000 Daltons covalently attached to (a) at least one
 CC chelating moiety selective for a monovalent or divalent metal ion (to
 CC act as a fluorescent indicator for the metal ion) and (b) at least one
 CC targeting peptide (to localise the compound to the inside of a cellular
 CC organelle). Preferably the polysaccharide is dextran, ficol, heparin,
 CC glycogen, amylopectin, mannan, inulin, starch, agarose or cellulose with
 CC a molecular weight less than 3,000,000. The metal ion is preferably Na⁺,
 CC K⁺, Li⁺, Ca²⁺ or Mg²⁺. The compounds are useful for analysing intra-
 CC cellular ion levels, especially Ca²⁺. The present sequence represents a
 CC carboxyl terminal sequence commonly shared by localisation peptides
 CC for the endoplasmic reticulum (ER).
 CC
 SQ Sequence 4 AA;
 Query Match 100.0%; Score 20; DB 19; Length 4;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 KDEL 4
 DB 1 KDEL 4
 RESULT 15
 AAM51437
 ID AAM51437 standard; peptide; 4 AA.
 AC AAM51437;
 XX
 XX 02-SEP-1998 (first entry)
 DT
 XX
 DE C-terminal sequence of endoplasmic reticulum retained proteins.
 XX
 KW Endoplasmic reticulum; carboxy terminus; Golgi apparatus;
 KW glycosyltransferase; membrane anchor; oligosaccharide synthesis.
 XX
 OS Synthetic.
 XX
 XX US5776772-A.
 PN
 XX 07-JUL-1998.
 PD
 XX 30-JAN-1996; 96US-0533865.
 PF
 XX 27-MAY-1992; 92US-0849045.
 PR 24-OCT-1989; 89US-0426577.
 PR 10-MAR-1994; 94US-0209604.
 PR 30-JAN-1996; 96US-0533865.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Adler B, Browne JK, Colley KJ, Paulson JG, Ujita-Jee E;
 PI Weinstein J;
 XX WPI; 1998-398046/34.
 DR
 XX Expression cassette for producing glycosyltransferase in secretable
 PT form - lacking membrane anchor and Golgi retention signal, used for
 PT synthesis of oligosaccharide(s)

XX Disclosure; Column 3; 16pp; English.

PS
XX
CC The invention relates to expression cassettes which contain a promoter
CC and a DNA sequence encoding a secreted glycosyltransferase that lacks
CC the membrane anchor and the Golgi retention signal. The glycosyl-
CC transferases are used in enzymatic oligosaccharide synthesis.
CC They are produced in a soluble form that is secreted into the culture
CC medium without loss of enzymatic activity, resulting in increased
CC production and simpler recovery (free of any membrane-bound glycosyl-
CC transferase), compared with extraction from mammalian tissue. The
CC present sequence represents a carboxy-terminal sequence sufficient for
CC retention of a protein in the endoplasmic reticulum.

XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 19; Length 4;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 1 KDEL 4

Search completed: August 26, 2003, 16:04:29
Job time : 11.2832 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:05:48 ; Search time 1.41593 Seconds
(without alignments)
119.528 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 KDEL 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	1	US-08-328-961-6 Sequence 6, Appl1
2	20	100.0	4	1	US-07-872-673B-5 Sequence 5, Appl1
3	20	100.0	4	1	US-08-405-615-16 Sequence 16, Appl1
4	20	100.0	4	1	US-08-331-398A-51 Sequence 51, Appl1
5	20	100.0	4	1	US-08-462-397-6 Sequence 6, Appl1
6	20	100.0	4	1	US-08-270-314-3 Sequence 3, Appl1
7	20	100.0	4	1	US-08-406-192-28 Sequence 28, Appl1
8	20	100.0	4	1	US-08-082-269D-7 Sequence 7, Appl1
9	20	100.0	4	1	US-08-593-865-1 Sequence 1, Appl1
10	20	100.0	4	2	US-08-461-234-16 Sequence 16, Appl1
11	20	100.0	4	2	US-08-480-190-152 Sequence 152, App
12	20	100.0	4	2	US-08-545-151-28 Sequence 28, Appl1
13	20	100.0	4	2	US-08-373-190-17 Sequence 17, Appl1
14	20	100.0	4	2	US-08-463-480-16 Sequence 16, Appl1
15	20	100.0	4	2	US-08-470-566B-51 Sequence 51, Appl1
16	20	100.0	4	2	US-08-488-379-152 Sequence 152, App
17	20	100.0	4	2	US-08-821-840-5 Sequence 5, Appl1
18	20	100.0	4	2	US-08-713-928B-15 Sequence 15, Appl1
19	20	100.0	4	2	US-08-407-900B-6 Sequence 6, Appl1
20	20	100.0	4	2	US-08-419-075-28 Sequence 28, Appl1
21	20	100.0	4	2	US-08-438-190A-17 Sequence 17, Appl1
22	20	100.0	4	2	US-08-923-536A-2 Sequence 2, Appl1
23	20	100.0	4	2	US-08-491-988-29 Sequence 29, Appl1
24	20	100.0	4	2	US-08-665-202-29 Sequence 29, Appl1
25	20	100.0	4	2	US-08-809-668-9 Sequence 9, Appl1
26	20	100.0	4	2	US-08-331-397B-51 Sequence 51, Appl1
27	20	100.0	4	2	US-08-759-804A-51 Sequence 51, Appl1

28	20	100.0	4	2	US-08-818-253-49 Sequence 49, Appl1
29	20	100.0	4	3	US-08-350-215-17 Sequence 17, Appl1
30	20	100.0	4	3	US-08-722-258-62 Sequence 62, Appl1
31	20	100.0	4	3	US-08-718-904-42 Sequence 42, Appl1
32	20	100.0	4	3	US-08-782-480-44 Sequence 44, Appl1
33	20	100.0	4	3	US-08-287-145A-17 Sequence 17, Appl1
34	20	100.0	4	3	US-09-397-951-9 Sequence 9, Appl1
35	20	100.0	4	3	US-08-776-271-7 Sequence 7, Appl1
36	20	100.0	4	3	US-09-047-148-12 Sequence 12, Appl1
37	20	100.0	4	3	US-09-094-359-17 Sequence 17, Appl1
38	20	100.0	4	3	US-09-112-063-37 Sequence 37, Appl1
39	20	100.0	4	3	US-08-769-333F-28 Sequence 28, Appl1
40	20	100.0	4	3	US-09-215-035-7 Sequence 7, Appl1
41	20	100.0	4	3	US-09-124-671-37 Sequence 37, Appl1
42	20	100.0	4	3	US-09-169-015-38 Sequence 38, Appl1
43	20	100.0	4	3	US-09-169-425C-23 Sequence 23, Appl1
44	20	100.0	4	3	US-09-191-852-1 Sequence 1, Appl1
45	20	100.0	4	3	US-08-818-252-49 Sequence 49, Appl1

ALIGNMENTS

RESULT 1
US-08-328-961-6
; Sequence 6, Application US/08328961
; Patent No. 5501975
; GENERAL INFORMATION:
; APPLICANT: Chaudhuri, Bhattacharya
; APPLICANT: Stephan, Christine
; APPLICANT: Seebach, Peter
; APPLICANT: Reizman, Howard
; TITLE OF INVENTION: No. 5501975el DNA Molecules and Hosts
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,961
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/989,260
; FILING DATE: 11-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 4-18885/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 1..4
; OTHER INFORMATION: /note="ER retention signal KDEL"
US-08-328-961-6
Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 2

US-07-872-673B-5
; Sequence 5, Application US/07872673B
; Patent No. 5578466

; GENERAL INFORMATION:

APPLICANT: Toshiya HAYANO, Setsuko KATO, No. 5578466hiro TAKAHASHI, and Masanori

TITLE OF INVENTION: Co-expression System of Protein Disulfide Isomerase Gene an

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS: ADDRESS: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS & CUSHMAN

STREET: 130 WATER STREET

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM: MEDIUM TYPE: Type 2DD, 3.50 inch, 720 KB

COMPUTER: Apple Macintosh SE

OPERATING SYSTEM: Apple DOS

SOFTWARE: Microsoft Word Version 4.0

CURRENT APPLICATION DATA: APPLICATION NUMBER: US/07/872,673B

FILING DATE: 19920417

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER: Japanese Patent Application No. 5578466, 114074/91 and 31160

FILING DATE: 18-APR-1991 and 30-OCT-1991

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 4

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: C-terminal fragment

FEATURE:

NAME/KEY: ER retention signal

LOCATION: C-terminus

OTHER INFORMATION: located at the C-terminus of rat PDI

PUBLICATION INFORMATION:

AUTHORS: Edman, Jeffrey C., Ellis, Ieland, Blacher, Russell W., Roth, Richa

TITLE: Sequence of protein disulphide isomerase and implications of its rel

Patent No. 5578466

JOURNAL: Nature

VOLUME: 317

PAGES: 267-270

DATE: 19-Sep-1985

RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 486 to 489

US-07-872-673B-5

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 2.5e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
|||
Db 1 KDEL 4

APPLICANT: Pastan, Ira

APPLICANT: Fitzgerald, David J.

TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with

TITLE OF INVENTION: Increased Activity

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS: ADDRESS: Ellen L. Weber

STREET: One Market Plaza, Stewart Tower, Suite 2000

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM: MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA: APPLICATION NUMBER: US/08/405,615

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER: US/07/901,709

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Ellen L.

REGISTRATION NUMBER: 32,762

REFERENCE/DOCKET NUMBER: 15280-36

TELEPHONE: 415-543-9600

TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 4 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-405-615-16

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 2.5e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 4

US-08-331-398A-51

; Sequence 51, Application US/08331398A

; Patent No. 5608039

; GENERAL INFORMATION:

APPLICANT: Pastan, Ira

APPLICANT: Williamson, Mark

APPLICANT: Fitzgerald, David

APPLICANT: Brinkmann, Ulrich

APPLICANT: Pai, Lee

TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins

TITLE OF INVENTION: and Their Uses (as amended)

NUMBER OF SEQUENCES: 68

CORRESPONDENCE ADDRESS: ADDRESS: Townsend and Townsend and Crew

STREET: One Market Plaza, Stewart Street Plaza

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105-1492

COMPUTER READABLE FORM: MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/331.398A
FILING DATE: 28-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/767,331
FILING DATE: 30-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/596,289
FILING DATE: 12-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Hunter, Tom
REGISTRATION NUMBER: 38,498
REFERENCE/DOCKET NUMBER: 015280-126110US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-331-398A-51

Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
Db 1 KDEL 4

RESULT 5
US-08-462-397-6
Sequence 6, Application US/08462397
Patent No. 5618690
GENERAL INFORMATION:
APPLICANT: Chaudhuri, Bhahatosh
APPLICANT: Stephan, Christine
APPLICANT: Seebach, Peter
APPLICANT: Reizman, Howard
TITLE OF INVENTION: No. 5618690e1 DNA Molecules and Hosts
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,397
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/989,260
FILING DATE: 11-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: 4-18885/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Domain
LOCATION: 1..4
OTHER INFORMATION: /note= "ER retention signal KDEL"
US-08-462-397-6

Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
Db 1 KDEL 4

RESULT 6
US-08-270-314-3
Sequence 3, Application US/08270314
Patent No. 568388
GENERAL INFORMATION:
APPLICANT: CAMPBELL, Anthony K.
TITLE OF INVENTION: MODIFIED BIOLUMINESCENT PROTEINS AND
TITLE OF INVENTION: THEIR USE
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: YOUNG & THOMPSON
STREET: 745 South 23rd Street
CITY: Arlington
STATE: VA
COUNTRY: US
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/270,314
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 89 16806.6
FILING DATE: 22-JUL-1989
APPLICATION NUMBER: PCT/GB90/01131
FILING DATE: 23-JUL-1990
APPLICATION NUMBER: US 07/820,867
FILING DATE: 22-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 521-2297
TELEFAX: (703) 685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-270-314-3

Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4

Db 1 KDEL 4

RESULT 7

US-08-406-192-28
Sequence 28, Application US/08406192

Patent No. 5739287
GENERAL INFORMATION:
APPLICANT: Wilbur, D. Scott
TITLE OF INVENTION: Prathare, Pradip M
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: WA 98101-2333
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/406,192
FILING DATE: 16-MAR-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/224,831
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Renzoni, George E
REGISTRATION NUMBER: 37,919
REFERENCE/DOCKET NUMBER: RECL18947
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
TELEX: 4938023
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-406-192-28

Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 1 KDEL 4

RESULT 8

US-08-082-269D-7
Sequence 7, Application US/08082269D

Patent No. 5773227
GENERAL INFORMATION:
APPLICANT: Kuhn, Michael
APPLICANT: Meyer, Tobias
TITLE OF INVENTION: Albritton, Nancy
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Molecular Probes, Inc.
STREET: 4849 Pitchford Avenue
CITY: Eugene
STATE: Oregon

COUNTRY: USA
ZIP: 97402-9144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch
COMPUTER: IBM
OPERATING SYSTEM: MS-DOS 6.2
SOFTWARE: Text Editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,269D
FILING DATE: 23-June-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Helmslein, Alegria J.
REGISTRATION NUMBER: 34,179
TELECOMMUNICATION INFORMATION:
TELEPHONE: (503)465-8300
TELEFAX: (503)344-6504
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 AMINO ACIDS
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: Peptide
HYPOTHETICAL: no
FRAGMENT TYPE:
PUBLICATION INFORMATION:
AUTHORS: Munro, Sean and Pelham Hugh, R.B.
TITLE: A C-Terminal Signal Prevents Secretion of Luminal ER Proteins
JOURNAL: Cell
VOLUME: 48
ISSUE: 13 March 1987
PAGES: 899-907
DATE: 1987
US-08-082-269D-7

Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 1 KDEL 4

RESULT 9

US-08-593-865-1
Sequence 1, Application US/08593865

Patent No. 5776772
GENERAL INFORMATION:
APPLICANT: Paulson, James C.
APPLICANT: Ujita-Lee, Bryn
APPLICANT: Colley, Karen J.
APPLICANT: Adler, Beverly
APPLICANT: Browne, Jeffrey K.
TITLE OF INVENTION: Weinstein, Jaeminder
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Poma, Smith, Lande & Rose
STREET: 2029 Century Park East, 38th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/593,865
FILING DATE: January 30, 1996

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 91/06635
FILING DATE: 16 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Oidenkamp, David J.
REGISTRATION NUMBER: 29,421
REFERENCE/DOCKET NUMBER: 117-033
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310)788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-593-865-1

Query Match 100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

RESULT 10
US-08-461-234-16
Sequence 16, Application US/08461234
Patent No. 5821238

GENERAL INFORMATION:
APPLICANT: Pasten, Ira H.
APPLICANT: Fitzgerald, David J.
TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with
TITLE OF INVENTION: Increased Activity
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Stuart Street Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,234
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,709
FILING DATE: 18-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,615
FILING DATE: 15-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 15280-36-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: peptide
HYPOTHETICAL: .NO
US-08-461-234-16

Query Match 100.0%; Score 20; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

RESULT 11
US-08-480-190-152
Sequence 152, Application US/08480190
Patent No. 5827516

GENERAL INFORMATION:
APPLICANT: Robert G. Urban
APPLICANT: Roman M. Chic
APPLICANT: Dario A. A. Vignali
APPLICANT: Mary L. Hedley
APPLICANT: Lawrence J. Stern
APPLICANT: Jack L. Strominger
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES
NUMBER OF SEQUENCES: 274
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S. A.
ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,190
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/077,255
FILING DATE: June 15, 1993
APPLICATION NUMBER: 07/925,460
FILING DATE: August 11, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00246/168001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 152:
SEQUENCE CHARACTERISTICS:
LENGTH: 4
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-480-190-152

Query Match 100.0%; Score 20; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

RESULT 12

US-08-545-151-28
; Sequence 28, Application US/08545151
; Patent No. 5840712
; GENERAL INFORMATION:
; APPLICANT: Morgan Jr., A. Charles
; APPLICANT: Wilbur, D. Scott
; TITLE OF INVENTION: Water Soluble Vitamin B12 Receptor
; TITLE OF INVENTION: Modulating Agents and Methods Related Thereto
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson & KindnessPRLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: WA 98101-2333
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,151
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04404
; FILING DATE: 07-APR-1995
; APPLICATION NUMBER: US 08/406,192
; FILING DATE: 16-MAR-1995
; APPLICATION NUMBER: US 08/406,194
; FILING DATE: 16-MAR-1995
; APPLICATION NUMBER: US 08/406,191
; FILING DATE: 16-MAR-1995
; APPLICATION NUMBER: US 08/224,831
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Renzoni, George E
; REGISTRATION NUMBER: 37,919
; REFERENCE/DOCKET NUMBER: RECL18878
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; TELEX: 4938023
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-545-151-28

Query Match 100.0%; Score 20; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 1 KDEL 4

RESULT 13
US-08-373-190-17
; Sequence 17, Application US/08373190
; Patent No. 5851829
; GENERAL INFORMATION:
; APPLICANT: MARASCO, WAYNE
; APPLICANT: HASELTINE, WILLIAM
; TITLE OF INVENTION: METHOD OF INTRACELLULAR BINDING OF PROTEINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Pasteo Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,190
; FILING DATE: 17-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06735
; FILING DATE: 16-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 41956-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: STRE UR 2002
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
US-08-373-190-17

Query Match 100.0%; Score 20; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 1 KDEL 4

RESULT 14
US-08-463-480-16
; Sequence 16, Application US/08463480
; Patent No. 5854044
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira H.
; APPLICANT: Fitzgerald, David J.
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with
; TITLE OF INVENTION: Increased Activity
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,480
; FILING DATE: 05-JUN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,709
FILING DATE: 18-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,615
FILING DATE: 15-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 15280-36-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
US-08-463-480-16

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 4;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

RESULT 15
US-08-470-566B-51
Sequence 51, Application US/08470566B
Patent No. 5872212
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kriesty
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5872212e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
US-08-470-566B-51

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 4;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

Search completed: August 26, 2003, 16:19:16
Job time : 2.41593 secs

This Page Blank (untested)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:02:18 ; Search time 1.41593 Seconds
(without alignments)
372.792 Million cell updates/sec

Title: US-09-696-872-37
Perfect score: 20
Sequence: 1 KDEL 4

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 497079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Published Applications_AA:*
2: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	8 US-08-765-244-16	Sequence 16, Appl
2	20	100.0	4	8 US-08-873-601-24	Sequence 24, Appl
3	20	100.0	4	9 US-09-157-748-30	Sequence 20, Appl
4	20	100.0	4	9 US-09-759-960-23	Sequence 23, Appl
5	20	100.0	4	9 US-09-854-122-2	Sequence 2, Appl
6	20	100.0	4	9 US-09-347-064-35	Sequence 35, Appl
7	20	100.0	4	9 US-09-333-527-2	Sequence 2, Appl
8	20	100.0	4	9 US-09-785-921A-16	Sequence 16, Appl
9	20	100.0	4	10 US-09-997-956-2	Sequence 2, Appl
10	20	100.0	4	10 US-09-781-804-2	Sequence 28, Appl
11	20	100.0	4	10 US-09-916-940-28	Sequence 8, Appl
12	20	100.0	4	10 US-09-480-236-8	Sequence 3, Appl
13	20	100.0	4	10 US-09-984-183-3	Sequence 22, Appl
14	20	100.0	4	10 US-09-999-745-22	Sequence 10, Appl
15	20	100.0	4	10 US-09-967-772-10	

16	20	100.0	4	10 US-09-554-000-49	Sequence 49, Appl
17	20	100.0	4	10 US-09-178-286-19	Sequence 19, Appl
18	20	100.0	4	10 US-09-792-630-77	Sequence 77, Appl
19	20	100.0	4	10 US-09-915-789A-10	Sequence 10, Appl
20	20	100.0	4	10 US-09-807-721-6	Sequence 6, Appl
21	20	100.0	4	10 US-09-825-803-3	Sequence 3, Appl
22	20	100.0	4	11 US-09-575-847-8	Sequence 8, Appl
23	20	100.0	4	11 US-09-991-209-97	Sequence 97, Appl
24	20	100.0	4	11 US-09-999-686-24	Sequence 24, Appl
25	20	100.0	4	11 US-09-906-393A-1	Sequence 1, Appl
26	20	100.0	4	11 US-09-861-257-19	Sequence 19, Appl
27	20	100.0	4	11 US-09-913-238-63	Sequence 63, Appl
28	20	100.0	4	11 US-09-462-713-13	Sequence 13, Appl
29	20	100.0	4	12 US-09-932-165-1508	Sequence 1508, Ap
30	20	100.0	4	12 US-10-177-725-157	Sequence 157, Ap
31	20	100.0	4	12 US-10-189-360-42	Sequence 42, Appl
32	20	100.0	4	12 US-10-226-877A-33	Sequence 33, Appl
33	20	100.0	4	12 US-10-252-732-8	Sequence 8, Appl
34	20	100.0	4	12 US-10-066-319-1	Sequence 1, Appl
35	20	100.0	4	12 US-10-133-973-63	Sequence 63, Appl
36	20	100.0	4	14 US-10-043-142-9	Sequence 9, Appl
37	20	100.0	4	14 US-10-154-801-3	Sequence 3, Appl
38	20	100.0	4	14 US-10-080-376-77	Sequence 77, Appl
39	20	100.0	4	14 US-10-061-395-25	Sequence 25, Appl
40	20	100.0	4	14 US-10-061-395-35	Sequence 35, Appl
41	20	100.0	4	15 US-10-096-339-24	Sequence 24, Appl
42	20	100.0	4	15 US-10-158-238-17	Sequence 17, Appl
43	20	100.0	4	15 US-10-127-427-2	Sequence 2, Appl
44	20	100.0	4	15 US-10-112-788-18	Sequence 18, Appl
45	20	100.0	4	15 US-10-053-355A-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-08-765-244-16
: Sequence 16, Application US/08765244
: Publication No. US20010008771A1
: GENERAL INFORMATION:
: APPLICANT: Seibel, Peter
: TITLE OF INVENTION: CHEMICAL PEPTIDE-NUCLEIC ACID
: TITLE OF INVENTION: FRAGMENT, PROCESS FOR PRODUCING THE SAME AND ITS USES FOR
: TITLE OF INVENTION: APPROPRIATELY INTRODUCING NUCLEIC ACIDS INTO CELL ORGANELLES
: TITLE OF INVENTION: AND CELLS
: FILE REFERENCE: 8484-0018-999
: CURRENT APPLICATION NUMBER: US/08/765,244
: PRIOR FILING DATE: 1997-10-30
: PRIOR APPLICATION NUMBER: PCT/DE95/00775
: PRIOR FILING DATE: 1995-06-11
: PRIOR APPLICATION NUMBER: DE P 44 21 079.5
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 16
: LENGTH: 4
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic sequence

US-08-765-244-16
Query Match 100.0%; Score 20; DB 8; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 2

```
US-08-873-601-24
; Sequence 24, Application US/08873601
; Publication No. US20020064798A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020064798A1an, Garry P.
; APPLICANT: Payan, Donald
; TITLE OF INVENTION: COMBINATORIAL ENZYMAIC COMPLEXES
; FILE REFERENCE: A-63915/DJB/RMS
; CURRENT APPLICATION NUMBER: US/08/873,601
; CURRENT FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: UNKNOWN
US-08-873-601-24

Query Match          100.0%; Score 20; DB 8; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KDEL 4
      ||||
Db      1 KDEL 4

RESULT 3
US-09-157-748-30
; Sequence 30, Application US/09157748
; Patent No. US20010003042A1
; GENERAL INFORMATION:
; APPLICANT: Lorens, James
; TITLE OF INVENTION: Multiparameter FACS Assays to Detect Alterations in
; TITLE OF INVENTION: Cell Cycle Regulation
; FILE REFERENCE: A66587/DJB/RMS
; CURRENT APPLICATION NUMBER: US/09/157,748
; CURRENT FILING DATE: 1998-09-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; JOURNAL: Royal Society London Transaction B
; PAGES: 1-1-
; DATE: 1992
; US-09-157-748-30

Query Match          100.0%; Score 20; DB 9; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KDEL 4
      ||||
Db      1 KDEL 4

RESULT 4
US-09-759-960-23
; Sequence 23, Application US/09759960
; Patent No. US20010006639A1
; GENERAL INFORMATION:
; APPLICANT: Urban, Robert G.
; APPLICANT: Chicz, Roman M.
; APPLICANT: Collins, Edward J.
; APPLICANT: Hedley, Mary Lynn
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES FROM THE HPV E7
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/759,960
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/169,425
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraezer, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 08191/004002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-543-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-759-960-23

Query Match          100.0%; Score 20; DB 9; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KDEL 4
      ||||
Db      1 KDEL 4

RESULT 5
US-09-854-122-2
; Sequence 2, Application US/09854122
; Patent No. US20020016980A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTE, RANDALL S.
; APPLICANT: SMITH, ROBERT
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA
; FILE REFERENCE: PHA-007.01
; CURRENT APPLICATION NUMBER: US/09/854,122
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-854-122-2

Query Match          100.0%; Score 20; DB 9; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KDEL 4
      ||||
Db      1 KDEL 4
```

Db 1 KDEL 4

RESULT 6

US-09-347-064-35
Sequence 35, Application US/09347064A
Patent No. US20020045208A1
GENERAL INFORMATION:
APPLICANT: Eck, Jürgen
APPLICANT: Schmidt, Arno
APPLICANT: Zinke, Holger
TITLE OF INVENTION: Recombinant Fusion Proteins Based on
TITLE OF INVENTION: Ribosome-Inactivating Proteins of the mistletoe Viscum
FILE REFERENCE: 09282-5
CURRENT APPLICATION NUMBER: US/09/347,064A
CURRENT FILING DATE: 1999-07-02
EARLIER APPLICATION NUMBER: PCT/EP98/00009
EARLIER FILING DATE: 1998-01-02
EARLIER APPLICATION NUMBER: EP 97 10 0012.0
EARLIER FILING DATE: 1997-01-02
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 4
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: Modulator
US-09-347-064-35

Query Match 100.0%; Score 20; DB 9; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

RESULT 7

US-09-333-527-2
Sequence 2, Application US/09333527
Patent No. US20020078472A1
GENERAL INFORMATION:
APPLICANT: Paul CHRISTOU, Eva STROGER, Rainer FISCHER, Carmen MARTIN-VAQUERO, Stef
TITLE OF INVENTION: METHODS AND MEANS FOR EXPRESSION OF MAMMALIAN POLYPEPTIDES
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,527
FILING DATE: Concurrently Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/089,322
FILING DATE: June 15, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Mary Anne Schofield
REGISTRATION NUMBER: 36,669
REFERENCE/DOCKET NUMBER: KL/JIC 202.1 - JEL
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4
TYPE: amino acid
TOPOLOGY: linear
US-09-333-527-2

Query Match 100.0%; Score 20; DB 9; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

RESULT 8

US-09-785-921A-16
Sequence 16, Application US/09785921A
Patent No. US20020094334A1
GENERAL INFORMATION:
APPLICANT: Keener, William K.
APPLICANT: Ward, Thomas E.
TITLE OF INVENTION: SELECTIVE DESTRUCTION OF CELLS INFECTED WITH HUMAN
FILE REFERENCE: LIT-PI-529
CURRENT APPLICATION NUMBER: US/09/785,921A
CURRENT FILING DATE: 2001-02-15
NUMBER OF SEQ ID NOS: 17
SEQ ID NO 16
LENGTH: 4
TYPE: PRT
ORGANISM: Homo sapiens
US-09-785-921A-16

Query Match 100.0%; Score 20; DB 9; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4
Db 1 KDEL 4

RESULT 9

US-09-997-956-2
Sequence 2, Application US/09997956
Patent No. US20020106714A1
GENERAL INFORMATION:
APPLICANT: Jalink, Kees
TITLE OF INVENTION: Membrane Molecule Indicator Compositions
FILE REFERENCE: P-NS 5045
CURRENT APPLICATION NUMBER: US/09/997,956
CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/250,679
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/256,559
PRIOR FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 4
TYPE: PRT
ORGANISM: Homo sapiens
US-09-997-956-2

Query Match 100.0%; Score 20; DB 10; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 10

US-09-781-804-2
; Sequence 2, Application US/09781804
; Patent No. US20020107189A1
; GENERAL INFORMATION:
; APPLICANT: ARID Gene Therapeutics, Inc.
; TITLE OF INVENTION: Regulation of Biological Events Using No. US20020107189A1 Comp
; FILE REFERENCE: 374 USD1
; CURRENT APPLICATION NUMBER: US/09/781,804
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: organelle binding domain
; NAME/KEY: DOMAIN
; LOCATION: (1)..(4)
; OTHER INFORMATION: organelle binding domain
US-09-781-804-2

Query Match 100.0%; Score 20; DB 10; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 11

US-09-916-940-28
; Sequence 28, Application US/09916940
; Patent No. US20020127564A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020127564A1a1, Garry P
; TITLE OF INVENTION: METHODS FOR SCREENING FOR TRANSDOMINANT INTRACELLULAR
; FILE REFERENCE: A-64260-6/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/916,940
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 09/727,715
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US 08/963,368
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: US 08/589,109
; PRIOR FILING DATE: 1996-01-23
; PRIOR APPLICATION NUMBER: US 08/589,911
; PRIOR FILING DATE: 1996-01-23
; PRIOR APPLICATION NUMBER: US 08/789,333
; PRIOR FILING DATE: 1997-01-23
; PRIOR APPLICATION NUMBER: US 08/787,738
; PRIOR FILING DATE: 1997-01-23
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: endoplasmic
; OTHER INFORMATION: reticulum sequence.
US-09-916-940-28

Query Match

Best Local Similarity 100.0%; Score 20; DB 10; Length 4;

Mismatches 100.0%; Pred. No. 4.4e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 12

US-09-480-236-8
; Sequence 8, Application US/09480236
; Patent No. US20020142000A1
; GENERAL INFORMATION:
; APPLICANT: Digan, Mary Ellen
; APPLICANT: Lake, Philip
; APPLICANT: Wright, Richard M.
; TITLE OF INVENTION: Anti-CD3 Immunotoxins and Therapeutic Uses Therefor
; FILE REFERENCE: CGC 4-31157A/USN
; CURRENT APPLICATION NUMBER: US/09/480,236
; CURRENT FILING DATE: 2000-01-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PE peptide
US-09-480-236-8

Query Match 100.0%; Score 20; DB 10; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 13

US-09-984-183-3
; Sequence 3, Application US/09984183
; Patent No. US20020142983A1
; GENERAL INFORMATION:
; APPLICANT: AGRAMAL, BABITA
; APPLICANT: LONGENECKER, MICHAEL B.
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE
; FILE REFERENCE: 042881/0130
; CURRENT APPLICATION NUMBER: US/09/984,183
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/457,354
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/111,973
; PRIOR FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Retrograde
; OTHER INFORMATION: transport peptide
US-09-984-183-3

Query Match 100.0%; Score 20; DB 10; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 14

US-09-999-745-22
Sequence 22, Application US/09999745
Patent No. US20020157120A1
GENERAL INFORMATION:
APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: Tsien, Roger Y.
APPLICANT: Baird, Geoffrey
TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
FILE REFERENCE: REGEN1470-1
CURRENT APPLICATION NUMBER: US/09/999,745
PRIOR FILING DATE: 2001-10-23
PRIOR APPLICATION NUMBER: 09/316,920
NUMBER OF SEQ ID NOS: 67
SOFTWARE: PatentIn version 3.0
SEQ ID NO 22
LENGTH: 4
TYPE: PRT
ORGANISM: Homo sapiens
US-09-999-745-22

Query Match 100.0%; Score 20; DB 10; Length 4;
Best Local Similarity 100.0%; Pred. No. 4,4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
|||
Db 1 KDEL 4

RESULT 15

US-09-967-772-10
Sequence 10, Application US/09967772
Patent No. US20020164577A1
GENERAL INFORMATION:
APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSJEN, Roger
APPLICANT: GONZALEZ, Jesus
TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
FILE REFERENCE: REGEN1290-5
CURRENT APPLICATION NUMBER: US/09/967,772
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: US 09/459,956
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 08/765,860
PRIOR FILING DATE: 1996-12-19
PRIOR APPLICATION NUMBER: PCT/ US96/09652
PRIOR FILING DATE: 1996-06-06
PRIOR APPLICATION NUMBER: US 08/481,977
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn version 3.0
SEQ ID NO 10
LENGTH: 4
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Protein localization sequence
US-09-967-772-10

Query Match 100.0%; Score 20; DB 10; Length 4;
Best Local Similarity 100.0%; Pred. No. 4,4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
|||
Db 1 KDEL 4

Search completed: August 26, 2003, 16:07:07
Job time : 2.41593 secs

This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 15:54:13 ; Search time 1.80531 Seconds
(without alignment)
213.080 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 KDEL 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	20	100.0	35	2	JX0200
2	20	100.0	38	2	A38869
3	20	100.0	38	2	C60657
4	20	100.0	38	2	B60657
5	20	100.0	41	2	B37188
6	20	100.0	46	2	S39239
7	20	100.0	50	2	S19624
8	20	100.0	50	2	AH2496
9	20	100.0	52	2	S19623
10	20	100.0	54	2	B96530
11	20	100.0	58	2	A35416
12	20	100.0	58	2	AD3018
13	20	100.0	60	2	AE2387
14	20	100.0	62	2	D63384
15	20	100.0	63	2	S11883
16	20	100.0	65	2	A12513
17	20	100.0	67	2	JU0399
18	20	100.0	68	2	G97737
19	20	100.0	70	2	AC3866
20	20	100.0	71	2	D81389
21	20	100.0	72	2	E81126
22	20	100.0	72	2	D84167
23	20	100.0	77	2	D69010
24	20	100.0	78	2	S69531
25	20	100.0	79	1	JN0246
26	20	100.0	79	2	B90352
27	20	100.0	79	2	D75326
28	20	100.0	79	2	AG2544
29	20	100.0	80	2	C69204

30	20	100.0	81	2	S77318	secretory protein
31	20	100.0	82	2	JU0765	rapid lysin III pr
32	20	100.0	82	2	D64329	hypothetical prote
33	20	100.0	83	2	B82891	hypothetical prote
34	20	100.0	83	2	B82887	hypothetical prote
35	20	100.0	85	2	I40586	hypothetical prote
36	20	100.0	85	2	T19536	probable involveme
37	20	100.0	86	1	WMECSR	replication contro
38	20	100.0	89	2	S05557	hypothetical prote
39	20	100.0	90	2	T04077	probable ribosomal
40	20	100.0	90	2	AC1030	hypothetical prote
41	20	100.0	91	2	E59100	hypothetical prote
42	20	100.0	92	2	H84264	hypothetical prote
43	20	100.0	92	2	H84296	hypothetical prote
44	20	100.0	93	2	T08319	conserved hypothet
45	20	100.0	93	2	AG2290	hypothetical prote

ALIGNMENTS

RESULT 1

JX0200
gummarin - *Gymnema sylvestre*
N:Alternate names: sweet-taste-suppressing peptide
C:Species: *Gymnema sylvestre*
C>Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 25-Oct-1996
C:Accession: JX0200
R:Kamei, K.; Takano, R.; Miyasaka, A.; Imoto, T.; Hara, S.
J. Biochem. 111, 109-112, 1992
A>Title: Amino acid sequence of sweet-taste-suppressing peptide (gummarin) from the leave
A:Reference number: JX0200; MUID:92299642; PMID:1607357
A:Accession: JX0200
A:Molecule type: Protein
A:Residues: 1-35 <KAM>
A:Experimental source: leaf
A:Comment: This protein suppresses sweet taste.
C:Keywords: pyrrolidonic acid; sweet taste
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 100.0%; Score 20; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 6 KDEL 9

RESULT 2

A38869
circumsporozoite protein - malaria parasite (*Plasmodium falciparum*) (isolate 427-5) (Frac
C:Species: *Plasmodium falciparum*
C>Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 09-Jun-2000
C:Accession: A38869
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.
Mol. Biochem. Parasitol. 37, 275-280, 1989
A>Title: Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T cell epi
A:Reference number: A60657; MUID:9011334; PMID:2481827
A:Accession: A38869
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-38 <LOC>
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 100.0%; Score 20; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
|||
Db 28 KDEL 31

```
RESULT 3
C60657
C:circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 7G8) (frag
C:Species: Plasmodium falciparum
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 09-Jun-2000
C:Accession: C60657
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.
Mol. Biochem. Parasitol. 37, 275-280, 1989
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep
A:Reference number: A60657; MUID:90114334; PMID:2481827
A:Accession: C60657
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-38 <LOC>
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match          100.0%; Score 20; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 28 KDEL 31

RESULT 4
B60657
C:circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate T9-98) (fra
C:Species: Plasmodium falciparum
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 09-Jun-2000
C:Accession: B60657
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.
Mol. Biochem. Parasitol. 37, 275-280, 1989
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep
A:Reference number: A60657; MUID:90114334; PMID:2481827
A:Accession: B60657
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-38 <LOC>
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match          100.0%; Score 20; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 28 KDEL 31

RESULT 5
B37188
major body fluid allergen ABA-1 - common roundworm (fragment)
C:Species: Ascaris lumbricoles (common roundworm)
C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 30-Sep-1993
C:Accession: B37188
R:Christie, J.F.; Dunbar, B.; Davidson, I.; Kennedy, M.W.
Immunology 69, 596-602, 1990
A:Title: N-terminal amino acid sequence identity between a major allergen of Ascaris lum
A:Reference number: A37188; MUID:90243308; PMID:2335378
A:Accession: B37188
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-41 <CHR>

Query Match          100.0%; Score 20; DB 2; Length 41;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 21 KDEL 24
```

```
RESULT 6
S39239
C:cholera enterotoxin chain A2 - Vibrio cholerae (serotype 0139)
N:Alternate names: CTx2 cholera toxin chain A2
C:Species: Vibrio cholerae
A:Variety: serotype 0139
C:Date: 08-May-1995 #sequence_revision 26-Jul-1996 #text_change 16-Jul-1999
C:Accession: S39239; S39240
R:Liebman, M.; Holmgren, U.
Submitted to the EMBL Data Library, November 1993
A:Description: Structure and arrangement of the Cholera toxin genes in Vibrio Cholera 01:
A:Reference number: S39238
A:Accession: S39239
A:Molecule type: DNA
A:Residues: 1-46 <LEB>
A:Cross-references: EMBL:X76390; NID:q433856; PIDN:CAAS3974.1; PID:q433858; EMBL:X76391;
A:Experimental source: strain 4260B; serotype 0139
C:Superfamily: heat-labile enterotoxin chain A

Query Match          100.0%; Score 20; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 43 KDEL 46

RESULT 7
S19624
ornatin B - leech (Placobdella ornata)
C:Species: Placobdella ornata
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S19624
R:Mazur, P.; Henzel, W.J.; Seymour, J.L.; Lazarus, R.A.
Eur. J. Biochem. 202, 1073-1082, 1991
A:Title: Ornatin: potent glycoprotein IIb-IIIa antagonists and platelet aggregation inh
A:Reference number: S19566; MUID:92111479; PMID:1765068
A:Accession: S19624
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-50 <MAZ>

Query Match          100.0%; Score 20; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 7 KDEL 10

RESULT 8
AH2496
transposase asr7152 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2496
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,
Nakazaki, N.; Shikimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2496
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-50 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA878236.1; PID:g17135690; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
```

A:Gene: asr7152
A:Genome: plasmid

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 50;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
|||
Db 8 KDEL 11

RESULT 9
S19623

ornatin C - leech (Placobdella ornata)

C:Species: Placobdella ornata
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997

C:Accession: S19623
R:Maizur, P.; Henzel, W.J.; Seymour, J.L.; Lazarus, R.A.

Eur. J. Biochem. 202, 1073-1082, 1991

A:Title: Ornatin: potent glycoprotein IID-IIIa antagonists and platelet aggregation inhibitors
A:Reference number: S19566; MUID:92111479; PMID:1765068

A:Accession: S19623

A:Status: preliminary
A:Molecule type: protein

A:Residues: 1-52 <MAZ>

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 52;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
|||
Db 7 KDEL 10

RESULT 10
E96530

hypochemical protein FljF21.16 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: E96530

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A6141; MUID:21016719; PMID:11130712

A:Accession: E96530

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-54 <STO>

A:Cross-references: GB:AE005173; NID:95430759; PIDN:AA43159.1; GSPDB:GN00141

C:Genetics:
A:Gene: FljF21.16
A:Map position: 1

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 54;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
|||
Db 47 KDEL 50

RESULT 11
A35416

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - bovine (fragments)

C:Species: Bos primigenius taurus (cattle)

C:Date: 18-Jan-1991 #sequence_revision 18-Jan-1991 #text_change 23-Jun-1993

C:Accession: A35416

R:Austin, C.A.; Barot, H.A.; Margerison, E.E.C.; Turcatti, G.; Winfield, P.; Hayes, M.V. Biochem. Biophys. Res. Commun. 170, 763-768, 1990

A:Title: Structure and partial amino acid sequence of calf thymus DNA topoisomerase II: c

A:Reference number: A35416; MUID:90343796; PMID:1696476

A:Accession: A35416

A:Status: preliminary
A:Molecule type: protein

A:Residues: 1-58 <AUS>

C:Keywords: isomerase

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
|||
Db 55 KDEL 58

RESULT 12
AD3018

conserved hypochemical protein Atu3752 [imported] - Agrobacterium tumefaciens (strain C58)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: AD3018

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L. erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krepsan, W.; Perry, M.; Gordon-Kamm, E ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AD3018

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-58 <KUR>

A:Cross-references: GB:AE008689; PIDN:AA144562.1; PID:917742178; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:
A:Gene: Atu3752

A:Map position: linear chromosome

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 58;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
|||
Db 46 KDEL 49

RESULT 13
AE2387

hypochemical protein asr4653 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AE2387
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE2387

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-60 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA876352.1; PID:917133790; GSPDB:GN00179

A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: asr4653

Query Match 100.0%; Score 20; DB 2; Length 60;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
 ||||
 DB 52 KDEL 55

RESULT 14

D69384
 hypothetical protein AF1076 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: D69384
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Author: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artiaeh, P.; Kaane, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
 A:Reference number: A69250; WUID:98049343; PMID:9389475
 A:Accession: D69384
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-62 <KDE>
 A:Cross-references: GB:AE001028; GB:AE000782; NID:g2689351; PIDN:AAB90163.1; PID:g264951

Query Match 100.0%; Score 20; DB 2; Length 62;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
 ||||
 DB 8 KDEL 11

RESULT 15

S11883
 copB protein - Escherichia coli plasmid ColV2-K94
 C:Species: Escherichia coli
 C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 30-May-1997
 C:Accession: S11883
 R:Banerjee, A.; Weber, P.C.; Palchaudhuri, S.
 Mol. Gen. Genet. 220, 320-324, 1990
 A>Title: Comparison of the CopB systems of plasmids R1 and ColV2-K94: a single base alte
 A:Reference number: S11883; WUID:90220506; PMID:2325626
 A:Accession: S11883
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-63 <MOL>
 C:Genetics:
 A:Gene: copB
 A:Genome: plasmid ColV2-K94
 C:Superfamily: repA2 protein

Query Match 100.0%; Score 20; DB 2; Length 63;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
 ||||
 DB 53 KDEL 56

Search completed: August 26, 2003, 16:05:26
 Job time : 6.80531 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:01:13 ; Search time 0.9915 Seconds
(without alignments)
189.786 Million cell updates/sec

Title: US-09-696-872-37
Sequence: 1 KDEL 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	28	1	ORND_PLAOR
2	20	100.0	35	1	GUR_GYMSY
3	20	100.0	37	1	VPV_HV128
4	20	100.0	50	1	ORNE_PLAOR
5	20	100.0	52	1	ORNC_PLAOR
6	20	100.0	62	1	YA76_ARCFU
7	20	100.0	67	1	Y00A_BPT4
8	20	100.0	76	1	Y066_HALN1
9	20	100.0	78	1	HOLI_BPH1
10	20	100.0	78	1	SI0D_HUMAN
11	20	100.0	81	1	SECE_SYNY3
12	20	100.0	82	1	VR3_BPT4
13	20	100.0	82	1	Y235_METUA
14	20	100.0	86	1	CPB3_ECOLI
15	20	100.0	89	1	DSBA_BPT4
16	20	100.0	97	1	Y04E_BPT4
17	20	100.0	101	1	ALP_LYMET
18	20	100.0	103	1	YC05_METUA
19	20	100.0	104	1	RECA_LEUME
20	20	100.0	105	1	Y349_METUA
21	20	100.0	105	1	YE95_MYCTU
22	20	100.0	109	1	RLAI_ARTSA
23	20	100.0	111	1	Y4E8_ENTRA
24	20	100.0	112	1	ANFG_AZONA
25	20	100.0	112	1	RBP3_SYNY3
26	20	100.0	112	1	Y474_METUA
27	20	100.0	114	1	THH1_ARATH
28	20	100.0	115	1	ANFG_CLOPA
29	20	100.0	116	1	ARSC_HAEIN
30	20	100.0	116	1	THIH_PAGES
31	20	100.0	116	1	Y472_METUA
32	20	100.0	118	1	THH2_TOBAC
33	20	100.0	118	1	THIH_RICCO

34	20	100.0	118	1	V118_ASPB7
35	20	100.0	120	1	RL31_PICMA
36	20	100.0	121	1	PEDB_METTH
37	20	100.0	123	1	RL7A_METKA
38	20	100.0	123	1	RL7A_METTH
39	20	100.0	123	1	RL7A_YEAST
40	20	100.0	126	1	RK12_CYAPA
41	20	100.0	127	1	UCR7_YEAST
42	20	100.0	128	1	RL7_GORGL
43	20	100.0	129	1	RS8_ECOLI
44	20	100.0	129	1	RS8_HAEIN
45	20	100.0	130	1	RS8_BUCAK

ALIGNMENTS

RESULT 1					
ORND_PLAOR	STANDARD;	PRT;	28 AA.		
AC P25513	01-MAY-1992 (Rel. 22, Created)				P18556 african swi
DT 01-MAY-1992 (Rel. 22, Last sequence update)					O65071 picea maria
DT 28-FEB-2003 (Rel. 41, Last annotation update)					O26774 methanobact
DE Ornatin D (Fragment).					O88v03 methanopyru
OS Placodella ornata (Turtle leech).					O26355 methanobact
OC Eukaryota; Metazoa; Amniota; Chordata; Hirdudinea; Hirudinea;					P19955 saccharomyc
OC Rhynchobdellida; Glossiphoniidae; Placodella.					P48124 cyanophora
OX NCBI_Taxid=6415;					P00128 saccharomyc
RN [1]					O8nt28 corynebacte
RP SEQUENCE.					P02361 escherichia
RX MEDLINE=92111479; PubMed=1765068;					P44377 haemophilus
RA Mazur P., Henzel W.J., Seymour J.L., Lazarus R.A.;					P46180 buchnera ap
RT "Ornatin: potent glycoprotein IIB-IIIA antagonists and platelet aggregation inhibitors from the leech Placodella ornata."					
RL Eur. J. Biochem. 202:1073-1082 (1991).					
CC -I- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF INGESTED BLOOD.					
CC -I- SIMILARITY: BELONGS TO THE ORNATIN FAMILY.					
DR InterPro: IPR002463; Ornatin.					
DR Pfam: PF02088; Ornatin; 1.					
KW Blood coagulation; Platelet; Cell adhesion.					
FT NON TER 28					
FT SEQUENCE 28 AA; 3361 MW; CFC38951F9137C2 CRC64;					
Query Match 100.0%; Score 20; DB 1; Length 28;					
Best Local Similarity 100.0%; Pred. No. 65;					
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY 1 KDEL 4					
DB 7 KDEL 10					
RESULT 2					
GUR_GYMSY	STANDARD;	PRT;	35 AA.		
ID GUR_GYMSY	01-MAY-1992 (Rel. 22, Created)				
AC P25810;	01-FEB-1984 (Rel. 28, Last sequence update)				
DT 01-FEB-1984 (Rel. 28, Last sequence update)	15-SEP-2003 (Rel. 42, Last annotation update)				
DE Gurmartin (Sweet-lacte-suppressing peptide).					
OS Gymnema sylvestre (Gurmar).					
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
OC Asteridae; Lamiales; Gentianales; Apocynaceae; Periplocoideae;					
OC Gymnema.					
OX NCBI_Taxid=4068;					
OX [1]					
RP SEQUENCE.					
RP TISSUE=leaf;					

```

RX MEDLINE=9229642; PubMed=1607357;
RA Kami K., Takano R., Miyasaka A., Imoto T., Hara S.;
RT "Amino acid sequence of sweet-taste-suppressing peptide (gummarin)
RL from the leaves of Gymnema sylvestre.";
RN J. Biochem. 111:109-112(1992).
[2]
RX STRUCTURE BY NMR.
RP MEDLINE=95307084; PubMed=7787425;
RA Arai K., Ichima R., Morikawa S., Miyasaka A., Imoto T., Yoshimura S.,
RN Aimoto S., Akasaka K.;
RT "Three-dimensional structure of gummarin, a sweet taste-suppressing
RL polypeptide.";
RN J. Biomol. NMR 5:297-305(1995).
[3]
RX STRUCTURE BY NMR.
RP MEDLINE=99421659; PubMed=10491100;
RA Fletcher J.I., Dingley A.J., Smith R., Connor M., Christie M.J.,
RN King G.F.;
RT "High-resolution solution structure of gummarin, a sweet-taste-
RL suppressing plant polypeptide.";
RN Eur. J. Biochem. 264:525-533(1999).
CC -!- FUNCTION: SUPPRESSES STRONGLY THE SWEET TASTE RESPONSES IN THE RAT
CC WITH HIGH SPECIFICITY TO SUCROSE, GLUCOSE, GLYCINE, AND SACCCHARIN.
CC THIS EFFECT IS REVERSIBLE, BUT COMPLETE RECOVERY OF THE SUPPRESSED
CC RESPONSES REQUIRED AT LEAST 3H. GUMMARIN SHOWED NO EFFECT OR ONLY
CC A VERY WEAK EFFECT ON THE SWEET TASTE SENSATION IN HUMANS.
CC PIR: JX0200; JX0200.
CC PDB: 1GUR; 01-AUG-96.
DR PDB: 1GUR; 27-AUG-96.
KW Taste-modifying protein; 3D-structure; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT DISULFID 3 18
FT DISULFID 10 23
FT DISULFID 17 33
FT TURN 12 15
FT STRAND 24 27
FT TURN 28 29
FT STRAND 30 32
SQ SEQUENCE 35 AA; 4232 MW; 661BD6FBAF504CB6 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   ||||
Db 6 KDEL 9

RESULT 3
VPU_HV128 STANDARD; PRT; 37 AA.
ID VPU_HV128
AC P08B07;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VPU protein (U ORF protein) (Fragment).
GN VPU
OS Human immunodeficiency virus type 1 (Z-84 isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
NCBI_TaxID=11681;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88281278; PubMed=3395517;
RA Youno J., Joseph S.F., Reitz M.S. Jr., Zagury D., Wong-Staal F.,
RN Gallo R.C.;
RT "Nucleotide sequence analysis of the env gene of a new Zairian
RL isolate of HIV-1.";
RT AIDS Res. Hum. Retroviruses 4:165-173(1988).
CC -!- FUNCTION: ACTS IN THE DEGRADATION OF CD4 IN THE ENDOPLASMIC
CC RETICULUM AND IN THE ENHANCEMENT OF VIRION RELEASE FROM THE PLASMA
CC MEMBRANE OF INFECTED CELLS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound.

```

```

CC -!- MISCELLANEOUS: THE Z-84 ISOLATE WAS TAKEN FROM A 54 YEAR-OLD
CC ZAIREAN MALE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL: J03653; AAA44683.1; -.
DR HSSP: P19554; IVPV.
DR HIV: J03653; VPU$JY1.
DR InterPro: IPR002094; Vpu.
DR Pfam: PF00558; Vpu; 1.
KW Transmembrane; AIDS.
FT NON_TER 1 1
SQ SEQUENCE 37 AA; 4131 MW; 00EA53CA27D09509 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   ||||
Db 17 KDEL 20

RESULT 4
ORNE_PLAOR STANDARD; PRT; 50 AA.
ID ORNE_PLAOR
AC P2514;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ornatin E.
OS Placobdella ornata (Turtle leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
OC Rhynchobdellida; Glossiphoniidae; Placobdella.
NCBI_TaxID=6415;
RN [1]
RP SEQUENCE.
RX MEDLINE=92111479; PubMed=1765068;
RA Mazur P., Henzel W.J., Seymour J.L., Lazarus R.A.;
RT "Ornatin: potent glycoprotein Iib-IIa antagonists and platelet
RT aggregation inhibitors from the leech Placobdella ornata.";
RL Eur. J. Biochem. 202:1073-1082(1991).
CC -!- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET
CC RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIA COMPLEX. MAY PREVENT
CC BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF
CC INGESTED BLOOD.
CC -!- SIMILARITY: BELONGS TO THE ORNATIN FAMILY.
CC PIR: S19624; S19624.
DR InterPro: IPR002463; Ornatin.
DR Pfam: PF02088; Ornatin; 1.
DR ProDom: PD012062; Ornatin; 1.
KW Blood coagulation; Platelet; Cell adhesion.
FT SITE 42 44 CELL ATTACHMENT SITE.
SQ SEQUENCE 50 AA; 5727 MW; 688408EB1E00192 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 1,2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   ||||
Db 7 KDEL 10

RESULT 5
ORNC_PLAOR STANDARD; PRT; 52 AA.
ID ORNC_PLAOR

```



```

AC P25512;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ornatin C.
OS Placobdella ornata (Turtle leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea;
OC Rhynchobdellida; Glossiphoniidae; Placobdella.
OX NCBI_TaxID=6415;
RN [1]
RP SEQUENCE.
RX MEDLINE=92111479; PubMed=1765068;
RA Maun P., Henzel W.J., Seymour J.L., Lazarus R.A.;
RT "Ornatins: potent glycoprotein iib-iiia antagonists and platelet
RT aggregation inhibitors from the leech Placobdella ornata.";
RL Eur. J. Biochem. 202:1073-1082(1991)
CC -1- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET
CC RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT
CC BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF
CC INGESTED BLOOD.
CC -1- SIMILARITY: BELONGS TO THE ORNATIN FAMILY.
DR PIR: S19623; S19623.
DR InterPro; IPR002463; Ornatin.
DR Pfam; PF02088; Ornatin; 1.
DR ProDom; PD012062; Ornatin; 1.
KW Blood coagulation; Platelet; Cell adhesion.
FT SITE 42 44
FT SEQUENCE 52 AA; 5845 MW; BA55CA7408EF4F09 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
DB 7 KDEL 10

RESULT 6
YA76_ARCFU STANDARD; PRT; 62 AA.
ID YA76_ARCFU
AC O29187;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1076.
GN AF1076.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
KLANK H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerkelaugh A.R., Graham D.E., Kyrle N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kitzner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Maeson T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001028; AAB90163.1; -.
DR PIR; D69384; D69384.
DR TIGR; AF1076; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 62 AA; 7385 MW; E2F541BA57A10477 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
DB 8 KDEL 11

RESULT 7
Y00A_BPT4 STANDARD; PRT; 67 AA.
ID Y00A_BPT4
AC P25186;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 8.1 kDa protein in rIIA-gp60 intergenic region.
GN Y00A OR RIIA.1.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90337270; PubMed=2379817;
RA Daegelen P., Brody E.;
RT "The rIIA gene of bacteriophage T4. I. Its DNA sequence and discovery
RT of a new open reading frame between genes 60 and rIIA.";
RL Genetics 125:237-248(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Kuter E., Arisaka F., Kunitawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52686; CA36910.1; -.
DR EMBL; AF158101; AAD42648.1; -.
DR PIR; JU0399; JU0399.
KW Hypothetical protein.
SQ SEQUENCE 67 AA; 8129 MW; B2DAA25C721A8753 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 67;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
DB 52 KDEL 55

RESULT 8
Y066_HALN1 STANDARD; PRT; 76 AA.
ID Y066_HALN1

```

```

AC Q9HSU8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Vng00066h.
GN VNG00066h.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxId=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Laskey S.R., Baliga N.S., Thorson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Lettner S., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddock D.G., Jablonki P.B., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Beck R.F., Pohl-Schroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1."
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RL -1- SIMILARITY: BELONGS TO THE UPF0175 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB004976; AAG18704.1; -
CC PIR; D84167; D84167.
CC InterPro; IPR005368; UPF0175.
CC Pfam; PF03683; UPF0175; 1.
CC DR Hypothetical protein; Complete proteome.
CC SEQUENCE 76 AA; 8577 MW; 408356DCFA6C87B CRC64;
SQ
Query Match 100.0%; Score 20; DB 1; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KDEL 4
DB 47 KDEL 50

```

```

RA Eposito D., Fitzmaurice W.P., Benjamin R.C., Goodman S.D.,
RA Waldman A.S., Socca J.J.;
RT "The complete nucleotide sequence of bacteriophage HPI DNA.";
RL Nucleic Acids Res. 24:2360-2368(1996).
CC -1- FUNCTION: INDUCES HOST CELL LYSIS BY CAUSING FORMATION OF PORES IN
CC THE INNER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U24159; AB09210.1; -
CC PIR; S69531; S69531.
CC Phage lysis protein; Transmembrane; Inner membrane.
CC TRANSMEM 8 24
CC SEQUENCE 78 AA; 8694 MW; 3FC0596BE0A4FED CRC64;
SQ
Query Match 100.0%; Score 20; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KDEL 4
DB 55 KDEL 58

```

```

RESULT 9
HOLI_BPHPI
ID HOLI_BPHPI STANDARD; PRT; 78 AA.
AC P51727;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HOLI.
GN HOLI.
OS Bacteriophage HPI.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX NCBI_TaxId=10690;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HPIc1;
RC MEDLINE=85128433; PubMed=6098523;
RA Benjamin R.C., Fitzmaurice W.P., Huang P.C., Socca J.J.;
RT "Nucleotide sequence of cloned DNA segments of the Haemophilus
RT influenzae bacteriophage HPIc1."
RL Gene 31:173-185(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=HPIc1;
RC MEDLINE=96279738; PubMed=8710508;

```

```

RESULT 10
SID_HUMAN
ID SID_HUMAN STANDARD; PRT; 78 AA.
AC P29377;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vitamin D-dependent calcium-binding protein, intestinal (CABP)
DE (Calbindin D9k).
GN CAB3 OR S100D OR CABP9K.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92304291; PubMed=1610358;
RA Howard A., Legon S., Spurr N.K., Walters J.R.I.;
RT "Molecular cloning and chromosomal assignment of human calbindin-
RT D9k."
RL Biochem. Biophys. Res. Commun. 185:663-669(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92354716; PubMed=1379540;
RA Jeung B.B., Kristinger J., Dann J.L., Jeung P.C.K.;
RT "Molecular cloning of the full-length cDNA encoding the human
RT calbindin-D9k."
RL FEBS Lett. 307:224-228(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94141916; PubMed=830886;
RA Jeung B.B., Jeung P.C.K., Kristinger J.;
RT "The human calbindin-D9k gene. Complete structure and implications on
RT steroid hormone regulation."
RL J. Mol. Biol. 235:1231-1238(1994).
CC -1- SIMILARITY: BELONGS TO THE S-100 FAMILY.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65869; CAA46699.1; -
DR EMBL; L13420; AAA35638.1; -
DR EMBL; L13420; AAA35637.1; -
DR PIR; JN0246; JN0246.
DR HSSP; P02632; ICBI.
DR Genew; HGNC:1436; CALB3.
DR MIM; 302020; -
DR GO; GO:0005509; F:calcium ion binding activity; TMS.
DR InterPro; IPR001751; CABP_S100.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR Pfam; PF01023; S_100; 1.
DR ProDom; PD003407; CABP_S100; 1.
DR ProDom; PD000012; EF-hand; 1.
DR PROSITE; PS00018; EF-HAND; 1.
DR PROSITE; PS00303; S100 CABP; 1.
DR Calcium-binding; Vitamin D; Acetylation.
KM INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT CA_BIND 17 30 EF-HAND 1 (LOW AFFINITY).
FT CA_BIND 57 68 EF-HAND 2 (HIGH AFFINITY).
FT CONFLICT 78 78 O -> S (IN REF. 3).
SQ SEQUENCE 78 AA; 8885 MW; 110FA5B92485E9F6 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 KDEL 4
DB 28 KDEL 31

RESULT 11
SECE_SYNY3
ID SECE_SYNY3 STANDARD; PRT; 81 AA.
AC P38382;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Preprotein translocase secE subunit.
GN SECE OR SCL3335.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93281410; PubMed=7685084;
RA Schmidt J., Subiramanian A.R.;
RT "Sequence of the cyanobacterial rRNA(w) gene in Synchocystis PCC
RT 6803: requirement of enzymatic 3' CCA attachment to the acceptor-
RT stem";
RL Nucleic Acids Res. 21:2519-2519(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Koreani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nanno K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions";
RL DNA Res. 3:109-116(1996).
RN [3]
CC -1- FUNCTION: ESSENTIAL FOR PROTEIN EXPORT.
CC -1- SUBCELLULAR LOCATION: Tail-anchored membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the secE/SEC61-gamma family.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X72627; -; NOT ANNOTATED_CDS.
DR EMBL; D90906; BAA17421.1; -
DR PIR; S77318; S77318.
DR InterPro; IPR001901; SECE.
DR InterPro; IPR005807; SECE_bac.
DR Pfam; PF00584; SECE; 1.
DR TIGRFAMs; TIGR00964; 3a0501806; 1.
DR PROSITE; PS01067; SECE_SEC61G; 1.
KM Protein transport; Translocation; Transmembrane; Complete proteome.
FT TRANSMEM 50 70 POTENTIAL.
SQ SEQUENCE 81 AA; 9134 MW; C3D3E0BA8100209F CRC64;

Query Match 100.0%; Score 20; DB 1; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 KDEL 4
DB 31 KDEL 34

RESULT 12
VR3_BPT4
ID VR3_BPT4 STANDARD; PRT; 82 AA.
AC P17309;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein rIII.
GN RIII OR 31.-1.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D;
RX MEDLINE=90301484; PubMed=2362813;
RA Prillipov A.G., Mesyanzhinov V.V., Aebi U., Kellenberger E.;
RT "Cloning and sequencing of bacteriophage T4 genes between map
RT positions 128.3-130.3";
RL Nucleic Acids Res. 18:3635-3635(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90332452; PubMed=2377483;
RA Raudonkluene A., Nivinskas R.;
RT "Nucleotide sequence of bacteriophage T4 gene 31 region";
RL Nucleic Acids Res. 18:4280-4280(1990).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=92267389; PubMed=1587487;
RA Raudonkluene A., Nivinskas R.;
RT "Gene rIII is the nearest downstream neighbour of bacteriophage T4
RT gene 31.";
RL Gene 114:85-90(1992).
RN [4]
RP SEQUENCE FROM N.A., AND MUTANTS.
RX MEDLINE=94063508; PubMed=8244025;
RA Raudonkluene A., Nivinskas R.;
RT "The sequences of gene rIII of bacteriophage T4 and its mutants";
RL Gene 134:135-136(1993).
RN [5]
RP SEQUENCE FROM N.A.
RA Kutter E., Aizsaka F., Kunisawa T., Tsugita A., Mosig G.,

```

RA Mesyanzhinov V., Ruger W., Stihham T., Thomas E.;
RL "Bacteriophage T4 genome analysis.";
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17657; CAA35652.1; -
DR EMBL; M37882; AAA32507.1; -
DR EMBL; X54536; CAA38406.1; -
DR EMBL; AF158101; AAD42650.1; -
DR PIR; J70765; J70765.
FT VARIANT 42 42
FT H -> R (IN R67 MUTANT; R-TYPE PLAQUE
FT MORPHOLOGY).
FT K -> E (IN RES40 AND RBB9 MUTANTS; R-TYPE
FT PLAQUE MORPHOLOGY).
FT KPLASARKAVHFVTLK -> NMLLHKEQFTLM
FT (IN REF. 1).
SQ SEQUENCE 82 AA; 9325 MW; CFE8DF45AB9CC2CD CRC64;
Query Match 100.0%; Score 20; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 57 KDEL 60

RESULT 13
ID Y235_METUA STANDARD; PRT; 82 AA.
AC 057687;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0235.
OS MJ0235.
GN Methanococcus jannaschii.
OC Archaeae; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,
RA Scott J.L., Geisheisen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utermbeck T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii."
RT Science 273:1058-1073 (1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67479; AAB88227.1; -
DR PIR; D64329; D64329.

```

```

DR TIGR; MJ0235; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 82 AA; 9899 MW; CB23BFED0A7F4E9F CRC64;
Query Match 100.0%; Score 20; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 36 KDEL 39

RESULT 14
ID CPB3_ECOLI STANDARD; PRT; 86 AA.
AC P03855;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Replication regulatory protein repA2 (copB protein).
GN REP22 OR COPB OR REPB OR CP0258.
OS Escherichia coli; and
OS Shigella flexneri.
OG Plasmid IncFII R1-19 (R1 dtd-19), and plasmid pCP301.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; PLASMID=IncFII R1-19 (R1 dtd-19);
RX MEDLINE=81172236; PubMed=6261081;
RA Stougaard P., Molin S., Nordstrom K., Hansen F.G.;
RT "The nucleotide sequence of the replication control region of the
RT resistance plasmid R1dtd-19."
RL Mol. Gen. Genet. 181:116-122 (1981).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S. flexneri; STRAIN=301 / Serotype 2a; PLASMID=pCP301;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao Y., Gao Y., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441 (2002).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE DETERMINATION OF COPY
CC NUMBER IN GENE REPLICATION. IT BINDS TO THE REPA PROMOTER THUS
CC INHIBITING THE SYNTHESIS OF THE MRNA FOR THE INITIATOR PROTEIN
CC REPA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF386526; AL72557.1; -
DR PIR; A04484; WMECSR.
KW Plasmid; Plasmid copy control; Transcription regulation; Repressor;
KW DNA-binding.
SQ SEQUENCE 86 AA; 10038 MW; F7AD85B2C211CBF8 CRC64;
Query Match 100.0%; Score 20; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 57 KDEL 60

```

Db 76 KDEL 79

RESULT 15

DSBA_BPT4 STANDARD; PRT; 89 AA.

AC P13320;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Double-stranded binding protein (dsDNA binding protein A).

GN DSBA OR RPB8.

OS Bacteriophage T4.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;

OC T4-like viruses.

OX NCBI_TaxId=10665;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=BK536;

RX MEDLINE=89386003; PubMed=2674900;

RA Hahn S., Rueger W.;

RT "Organization of the bacteriophage T4 genome between map positions

RT 150,745 and 145,824.";

RL Nucleic Acids Res. 17:6729-6729 (1989).

RN [2] SEQUENCE FROM N.A.

RA Kutter E., Arisaka F., Kunitawa T., Tsugita A., Mosig G.,

RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;

RT "Bacteriophage T4 genome analysis.";

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [3] CHARACTERIZATION.

RX MEDLINE=91203817; PubMed=2017138;

RA Ganz A., Kruse U., Rueger W.;

RT "Gene product dsbA of bacteriophage T4 binds to late promoters and

RT enhances late transcription.";

RL Mol. Gen. Genet. 225:427-433 (1991).

CC -! FUNCTION: BINDS DOUBLE-STRANDED DNA; INTERACTS PREFERENTIALLY WITH

CC T4 LATE PROMOTER REGIONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X15818; CAA33813.1; -.

DR EMBL; AF158101; AAD42567.1; -.

DR PIR; S05557; S05557.

KW DNA replication; DNA-binding.

SQ SEQUENCE 89 AA; 10379 MW; 0152649899E09319 CRC64;

Query Match

Best Local Similarity 100.0%; Score 20; DB 1; Length 89;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4

Db 47 KDEL 50

Search completed: August 26, 2003, 16:06:21

Job time : 2.99115 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:04:38 ; Search time 4.31858 Seconds
(without alignments)
239.016 Million cell updates/sec

Title: US-09-696-872-37
Perfect score: 20
Sequence: 1 KDEL 4

Scoring table: BLOSUM62
Gapox 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_todent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	21	11 Q9ESX0	Q9ESX0 mus musculu
2	20	100.0	35	10 Q9S8D1	Q9S8D1 gymema syl
3	20	100.0	40	4 Q13833	Q13833 homo sapien
4	20	100.0	40	4 Q13832	Q13832 homo sapien
5	20	100.0	41	2 O06527	O06527 lactobacill
6	20	100.0	41	11 Q9ECR6	Q9ECR6 mus musculu
7	20	100.0	41	16 Q8DRV6	Q8DRV6 streptococc
8	20	100.0	42	5 Q9NG47	Q9NG47 aedes albp
9	20	100.0	46	2 Q57372	Q57372 vibrio chol
10	20	100.0	48	4 Q9UDJ7	Q9UDJ7 homo sapien
11	20	100.0	49	16 Q8EB07	Q8EB07 shewanella
12	20	100.0	50	2 Q50051	Q50051 mycobacteri
13	20	100.0	50	5 Q8ST38	Q8ST38 anartia fat
14	20	100.0	50	5 Q9SVF9	Q9SVF9 dryas iulia
15	20	100.0	50	16 Q8YKX6	Q8YKX6 anabaena sp
16	20	100.0	51	5 Q8T5E2	Q8T5E2 anartia lyc

17	20	100.0	51	5 Q95VG2	Q95VG2 eueides ali
18	20	100.0	51	5 Q95VG0	Q95VG0 eueides vib
19	20	100.0	51	5 Q8T5D0	Q8T5D0 anartia fat
20	20	100.0	52	5 Q8T5D9	Q8T5D9 anartia ama
21	20	100.0	52	5 Q8T5D1	Q8T5D1 anartia fat
22	20	100.0	52	10 Q9SFX4	Q9SFX4 picea abies
23	20	100.0	52	10 Q65057	Q65057 picea maria
24	20	100.0	53	5 Q8T5E1	Q8T5E1 anartia chr
25	20	100.0	53	7 Q30456	Q30456 equus cabal
26	20	100.0	53	7 Q30476	Q30476 equus cabal
27	20	100.0	53	7 Q30471	Q30471 equus cabal
28	20	100.0	53	7 Q30470	Q30470 equus cabal
29	20	100.0	53	7 Q9BCX3	Q9BCX3 equus cabal
30	20	100.0	53	10 Q9SPX3	Q9SPX3 picea abies
31	20	100.0	53	16 Q8KER6	Q8KER6 chlorobium
32	20	100.0	54	10 Q9XIA7	Q9XIA7 arabidopsis
33	20	100.0	57	11 Q9ES41	Q9ES41 rattus norv
34	20	100.0	58	16 Q8U9H6	Q8U9H6 agrobacteri
35	20	100.0	59	5 Q8T5E0	Q8T5E0 anartia chr
36	20	100.0	60	16 Q8YNB4	Q8YNB4 anabaena sp
37	20	100.0	60	17 Q97C12	Q97C12 thermoplasm
38	20	100.0	61	11 Q88367	Q88367 mus musculu
39	20	100.0	61	16 Q9ADC0	Q9ADC0 streptomyc
40	20	100.0	62	13 Q91801	Q91801 xenopus lae
41	20	100.0	62	13 Q8JH23	Q8JH23 ostracion s
42	20	100.0	63	10 Q9AX34	Q9AX34 oryza sativ
43	20	100.0	64	4 Q96GN8	Q96GN8 homo sapien
44	20	100.0	65	5 Q8T5D2	Q8T5D2 anartia fat
45	20	100.0	65	5 Q8T5D8	Q8T5D8 anartia ama

ALIGNMENTS

RESULT 1	Q9ESX0	PRELIMINARY;	PRT;	21 AA.
ID	Q9ESX0			
AC	Q9ESX0;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Gephyrin (Fragment).			
GN	GPHN OR GEPHYRIN.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
NC	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20420367; PubMed=10963686;			
RA	Ramming M., Kins S., Werner N., Hermann A., Betz H., Kirsch J.			
RT	"Diversity and phylogeny of gephyrin: tissue-specific splice variants, gene structure, and sequence similarities to molluscan cofactor-synthesizing and cytoskeleton-associated proteins."			
RT	Proc. Natl. Acad. Sci. U.S.A. 97:10266-10271 (2000).			
DR	EMBL: AJ278768; CAC06105.1; -			
DR	MGP; MGI:109602; Gphn.			
FT	NON_TER			
FT	NON_TER			
SO	SEQUENCE	21 AA; 2627 MW; 0820F760BC776F9A CRC64;		
Query Match		100.0%; Score 20; DB 11; Length 21;		
Best local Similarity		100.0%; Pred. No. 3; e+02;		
Matches	4; Conservative	0; Mismatches	0; Indels	0; Gaps
Qy	1 KDEL 4			
Db	9 KDEL 12			
RESULT 2	Q9S8D1	PRELIMINARY;	PRT;	35 AA.
ID	Q9S8D1			

```

AC 0989D1;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE GURMARIN=SWEETNESS-suppressing polypeptide.
OS Gymnema sylvestre.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Gentianales; Apocynaceae; Periplocoidae; Gymnema.
OC NCBI_TaxID=4068;
RN [1]
RP MEDLINE=96068945; PubMed=8534991;
RA Oca M., Ariyoshi Y.;
RT "Location of the disulfide bonds of the sweetness-suppressing
RT polypeptide gurmarin.";
RL Bioact. Biotechnol. Biochem. 59:1956-1957(1995).
DR HSRP.P25810; IGUR.
SQ SEQUENCE 35 AA; 4233 MW; 661BCAE66E904CB6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 35;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 6 KDEL 9

RESULT 3
Q13833 PRELIMINARY; PRT; 40 AA.
AC 013833;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE B2 bradykinin receptor basal promoter, allele BP-58-T (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96209920; PubMed=8655154;
RA Braun A., Maier E., Kammerer S., Mueller B., Roscher A.A.;
RT "A novel sequence polymorphism in the promoter region of the human
RT bradykinin B2-receptor gene.";
RL Hum. Genet. 97:688-689(1996).
DR EMBL; X91664; CAA62852.1; -.
KW Receptor.
FT NON_TER 1 1
FT VARIANT 18 19 IT -> XS.
FT NON_TER 40 40
SQ SEQUENCE 40 AA; 4152 MW; 1408B9AD371EB17F CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 40;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 23 KDEL 26

RESULT 4
Q13832 PRELIMINARY; PRT; 40 AA.
AC 013832;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE B2 bradykinin receptor basal promoter, allele BP-58-C (Fragment).
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96209920; PubMed=8655154;
RA Braun A., Maier E., Kammerer S., Mueller B., Roscher A.A.;
RT "A novel sequence polymorphism in the promoter region of the human
RT bradykinin B2-receptor gene.";
RL Hum. Genet. 97:688-689(1996).
DR EMBL; X91663; CAA62851.1; -.
KW Receptor.
FT NON_TER 1 1
FT VARIANT 18 19 TT -> XS.
FT NON_TER 40 40
SQ SEQUENCE 40 AA; 4140 MW; 3908B9AD371EF4A5 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 40;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 23 KDEL 26

RESULT 5
Q06527 PRELIMINARY; PRT; 41 AA.
AC 006527;
DT 01-JUL-1997 (TRENBLREL. 04, Created)
DT 01-JUL-1997 (TRENBLREL. 04, Last sequence update)
DE 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Cystathionine gamma-lyase homolog (Fragment).
OS Lactobacillus fermentum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OC NCBI_TaxID=1613;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BR11;
RX MEDLINE=97294473; PubMed=9150229;
RA Turner M.S., Timms P., Hafner L.M., Giffard P.M.;
RT "Identification and characterization of a basic cell surface-located
RT protein from Lactobacillus fermentum BR11.";
RL J. Bacteriol. 179:3310-3316(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=BR11;
RX MEDLINE=99194729;
RA Turner M.S., Woodberry T., Hafner L.M., Giffard P.M.;
RT "The bcpA locus of Lactobacillus fermentum BR11 encodes an L-cysteine
RT uptake system.";
RL J. Bacteriol. 181:2192-2198(1999).
DR EMBL; U97348; AAC45329.1; -.
DR InterPro; IPR000277; Cys_Met_Meta_PP.
DR Pfam; PF01053; Cys_Met_Meta_PP; 1.
KW Lyase.
FT NON_TER 1 1
FT SEQUENCE 41 AA; 4647 MW; 7875F22817910FB7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 41;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 13 KDEL 16

RESULT 6
Q0906 PRELIMINARY; PRT; 41 AA.
ID Q0906

```



```

AC Q9E0E6;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE General transcription factor II-1 repeat domain protein 1
(Fragment).
GN GTF2IRD1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv.
MEDLINE=21250999; PubMed=11352562;
RA Duxkin M.E., Keck-Waggoner C.L., Popescu N.C., Thorgeirsson S.S.;
RT "Integration of a c-myc Transgene Results in Disruption of the Mouse
Gtf2ird1 Gene, the Homologue of the Human GTF2IRD1 Gene Hemizyously
Deleted in Williams-Beuren Syndrome.";
RL Genomics 73:20-27(2001).
DR EMBL; AF257477; AAC44656.1; -.
DR MGI; MGI:1861942; Gtf2ird1.
FT NON TER 41
SQ SEQUENCE 41 AA; 4482 MW; D7E29ECAS70059DC CRC64;

Query Match 100.0%; Score 20; DB 11; Length 41;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 27 KDEL 30

RESULT 7
Q8DRV6 PRELIMINARY; PRT; 41 AA.
AC Q8DRV6;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN SMU.2105.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
CX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE015032; AAM59699.1; -.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 41 AA; 4849 MW; CD8F7DC062E6D60B CRC64;

Query Match 100.0%; Score 20; DB 16; Length 41;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 5 KDEL 8

RESULT 8
Q9NG47 PRELIMINARY; PRT; 42 AA.

```

```

AC Q9NG47;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Ribosomal S5 protein (Fragment).
OS Aedes albopictus (Forest day mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.
CX NCBI_TaxID=7160;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang F.F., Zhao T.T., Li D.D.;
RT "Differentially expressed genes of Aedes albopictus orally infected
with dengue-2 virus.";
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF263471; AAF73440.1; -.
DR InterPro; IPR000235; Ribosomal_S7.
DR Pfam; PF00177; Ribosomal_S7; 1.
DR ProDom; PD000817; Ribosomal_S7; 1.
FT NON TER 1
SQ SEQUENCE 42 AA; 4711 MW; CE1FB5123BE053A2 CRC64;

Query Match 100.0%; Score 20; DB 5; Length 42;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 31 KDEL 34

RESULT 9
O57372 PRELIMINARY; PRT; 46 AA.
AC O57372;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Cholera toxin A2.
GN CTX42.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
CX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4260B;
RX MEDLINE=94237453; PubMed=8181723;
RA LeDens M., Holmgren J.;
RT "Structure and arrangement of the cholera toxin genes in Vibrio
cholerae O139.";
RL FEMS Microbiol. Lett. 117:197-202(1994).
DR EMBL; X76391; CAA53975.1; -.
DR EMBL; X76390; CAA53974.1; -.
DR HSSP; P01555; IXTC.
DR InterPro; IPR001144; Enterotoxin_A.
DR InterPro; IPR000886; ER_target.
DR Pfam; PF01375; Enterotoxin_A; 1.
DR ProSite; PS00014; ER_TARGET; 1.
SQ SEQUENCE 46 AA; 5447 MW; 1B6085A02E8889D6 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 43 KDEL 46

RESULT 10
Q9UDJ7 PRELIMINARY; PRT; 48 AA.

```

AC Q9UDJ7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Nucleolar shuttle protein B-23-38 kDa major NOS-binding protein
DE (Fragmente).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=93300776; PubMed=8314759;
RA Adachi Y., Copeland T.D., Hatanaka M., Oroszlan S.;
RT "Nucleolar targeting signal of Rex protein of human T-cell leukemia
RT virus type I specifically binds to nucleolar shuttle protein B-23.",
RL J. Biol. Chem. 268:13930-13934(1993).
FT NON_TER 1
FT NON_CONS 18 19
FT NON_CONS 33 34
FT NON_TER 48 48
SQ SEQUENCE 48 AA; 4783 MW; 5D7D533BD36A9B CRC64;

Query Match 100.0%; Score 20; DB 4; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 5 KDEL 8

RESULT 11

Q8EB07 PRELIMINARY; PRT; 49 AA.
AC Q8EB07;
ID 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
CN S03729.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Reed T.D., Eiken J.A., Seshadri R., Ward N., Mehe B., Clayton R.A.,
RA Meyer R.T., Tsai A., Scott J., Beanan M., Binkac L., Dougherty S.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Klonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis".
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015806; AAN56713.1; -.
DR TIGR; SO3729; -.
KW Hypothetical protein, Complete proteome.
SQ SEQUENCE 49 AA; 5340 MW; 7ABE848EAD0FFA80 CRC64;

Query Match 100.0%; Score 20; DB 16; Length 49;
Best Local Similarity 100.0%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 31 KDEL 34

RESULT 12
Q50051 PRELIMINARY; PRT; 50 AA.
AC Q50051;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE U2266p.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U15182; AAA62966.1; -.
SQ SEQUENCE 50 AA; 5563 MW; 5B5A4419AD5DD6C2 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 8.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 41 KDEL 44

RESULT 13

Q8ST38 PRELIMINARY; PRT; 50 AA.
AC Q8ST38;
ID 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Triosephosphate isomerase (Fragment).
CN S03729.
OS Anartia fatima.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Papilionoidea; Nymphalidae; Nymphalinae; Anartia.
OX NCBI_TaxID=145883;
RN [1]
RP SEQUENCE FROM N.A.
RA Blum M.J., Bermingham E., Dasmahapatra K.;
RT "A Molecular Phylogeny of the Neotropical Butterfly Genus Anartia
RT (Lepidoptera: Nymphalidae)".
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY038684; AAM03332.1; -.
DR EMBL; AY038685; AAM03333.1; -.
DR InterPro; IPR000652; Triophos_1smrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_1smrse; 1.
KW isomerase.
FT NON_TER 1
FT NON_TER 50 50
SQ SEQUENCE 50 AA; 5450 MW; F985EA898F5A3BAF CRC64;

Query Match 100.0%; Score 20; DB 5; Length 50;
Best Local Similarity 100.0%; Pred. No. 8.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 2 KDEL 5

RESULT 14

Q95VF9

Search completed: August 26, 2003, 16:09:20
 Job time : 7.31858 secs

```

ID Q95VF9 PRELIMINARY; PRT; 50 AA.
AC Q95VF9;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Triosephosphate isomerase (Fragment).
GN TPI.
OS Dryas julia (julia butterfly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Papilionoidea; Nymphalidae; Heliconiinae; Dryas.
OX NCB1_TaxID=33453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STRI-B-293-Tpi-1;
RA Beltran M.S., Jiggins C.D., Bull V., McMillan O., Bermingham E.,
RA Mallet J., Linhares M.;
RT "Phylogenetic discordance at the species boundary: comparative gene
RL genealogies between Heliconius butterflies.";
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF413797; AAL08558.1; -.
DR InterPro: IPR006552; Triophos_ismrse.
DR Pfam: PF00121; TIM; 1.
DR ProDom: PD001005; Triophos_ismrse; 1.
KW isomerase.
FT NON_TER 1
FT NON_TER 50
SQ SEQUENCE 50 AA; 5394 MW; AF3E2F54436BAFF8 CRC64;

```

Query Match 100.0%; Score 20; DB 5; Length 50;
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
 ||||
 Db 3 KDEL 6

RESULT 15

```

ID Q8YKY6 PRELIMINARY; PRT; 50 AA.
AC Q8YKY6;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Transposase.
GN ASR7152.
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCB1_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kiehlid Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003600; BAB78236.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 50 AA; 5760 MW; BE163ACA7B5C489B CRC64;

```

Query Match 100.0%; Score 20; DB 16; Length 50;
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
 ||||
 Db 8 KDEL 11

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: August 26, 2003, 16:08:15 ; Search time 177.416 Seconds
(without alignments) 922.344 Million cell updates/sec

Title: US-09-696-872-37
Perfect score: 20
Sequence: 1 KDEL 4

Scoring table: BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09696872/runat_26082003_151137_3223/app_query.fasta_1.462
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09696872.OCGN_1.1_3608.0@runat_26082003_151137_3223 -NCPU=6 -ICPU=3
-NO MMAP -LARGEODDERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ets.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_ets.*
28: em_un.*

29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hcgo_hum.*
40: em_hcgo_mus.*
41: em_hcgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	12	6	BD174826
2	20	100.0	16	6	AX710937
3	20	100.0	16	6	BD001078
4	20	100.0	16	6	BD001507
5	20	100.0	17	6	AX215210
6	20	100.0	17	6	AX215211
7	20	100.0	17	6	AX215212
8	20	100.0	17	6	AX216059
9	20	100.0	19	6	A58382
10	20	100.0	20	6	AR160690
11	20	100.0	20	6	AR183981
12	20	100.0	20	6	AR221427
13	20	100.0	20	6	AR221428
14	20	100.0	20	6	AR221429
15	20	100.0	20	6	BD138303
16	20	100.0	20	6	BD138304
17	20	100.0	21	6	BD143817
18	20	100.0	21	6	I26155
19	20	100.0	21	6	I86398
20	20	100.0	22	6	AR135254
21	20	100.0	23	6	AX343349
22	20	100.0	23	6	AX343350
23	20	100.0	23	6	BD133175
24	20	100.0	23	6	BD133176
25	20	100.0	24	6	A10287
26	20	100.0	24	6	A10287
27	20	100.0	24	6	A10288
28	20	100.0	24	6	A10288
29	20	100.0	24	6	A16561
30	20	100.0	24	6	A16561
31	20	100.0	24	6	A16562
32	20	100.0	24	6	A16562
33	20	100.0	24	6	A33707
34	20	100.0	24	6	A33707
35	20	100.0	24	6	A33708
36	20	100.0	24	6	A33708
37	20	100.0	24	6	A82182
38	20	100.0	24	6	AR307631
39	20	100.0	24	6	AR307641
40	20	100.0	24	6	BD132633
41	20	100.0	24	6	BD174866
42	20	100.0	25	6	AR137362
43	20	100.0	25	6	AX037484
44	20	100.0	25	6	AX370858
45	20	100.0	25	6	BD075236

RESULT 1

ALIGNMENTS

BD174826
LOCUS BD174826 12 bp DNA linear PAT 18-MAR-2003
DEFINITION Antibody-producing transgenic plant.
ACCESSION BD174826
VERSION BD174826.1 GI:29120518
KEYWORDS JP 2002253262-A/12.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 12)
AUTHORS Matsumura,T., Yohoda,N., Ito,K., Kato,M., Sugimoto,C., Ueda,I.,
Ohashi,K. and Li,C.
TITLE Antibody-producing transgenic plant
JOURNAL Patent: JP 2002253262-A 12 10-SEP-2002;
KK SCIENCE TANAKA,HOKKAIDO GREEN BIO INSTITUTE,CHIHIRO
SUGIMOTO,ICHIRO UEDA
COMMENT OS Unknown
PN JP 2002253262-A/12
PD 10-SEP-2002
PF 05-MAR-2001 JP 2001060462
PI TAKESHI MATSUMURA,NORIKO YOHODA,KEIZO ITO,MIHOKO KATO,CHIHIRO
PI SUGIMOTO,
PI ICHIRO UEDA,KAZUHIRO OHASHI,CHENGYI LI
PC C12N15/09,A01H5/00,C07K16/08,C12N5/10,C12Q1/70,G01N33/569, PC
G01N33/577//
PC C12P21/08,C12N15/00,C12N5/00
CC Description of Unknown Organism:KDEL signal
FT Key Location/Qualifiers
FT source 1..12
FT Location/Qualifiers
FT 1..12
FT /organism='Unknown'.
FT /mol_type='genomic DNA'
FT /db_xref='taxon:32644'
BASE COUNT 5 a 1 c 3 g 3 t
ORIGIN

Alignment Scores:
Pred. No.: 117 Length: 12
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-696-872-37 (1-4) x BD174826 (1-12)

RESULT 2
LOCUS AX710937 16 bp RNA linear PAT 11-APR-2003
DEFINITION Sequence 237 from Patent EP1288296.
ACCESSION AX710937
VERSION AX710937.1 GI:29787318
KEYWORDS Human herpesvirus 5
SOURCE Human herpesvirus 5
ORGANISM Human herpesvirus 5
REFERENCE 1
AUTHORS Draper,K.G., McSwigen,J.A., Holecek,J.J., Dudycz,L.W.,
Macejak,D.G. and Mamone,J.A.
TITLE Method and reagent for inhibiting HBV viral replication
JOURNAL Patent: EP 1288296-A 237 05-MAR-2003;
RIBOZYME PHARMACEUTICALS, INC. (US)
FEATURES
SOURCE 1..16
/organism='Human herpesvirus 5'
/mol_type='genomic RNA'

BASE COUNT 7 a 1 c 5 g 3 t
ORIGIN /db_xref='taxon:10359'

Alignment Scores:
Pred. No.: 158 Length: 16
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-696-872-37 (1-4) x AX710937 (1-16)

RESULT 3
LOCUS BD001078 16 bp RNA linear PAT 31-JAN-2002
DEFINITION Method and reagent for inhibiting viral replication.
ACCESSION BD001078
VERSION BD001078.1 GI:18625637
KEYWORDS JP 2000342285-A/238.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 16)
AUTHORS Draper,K.G., Dudycz,L.W., McSwigen,J.A., Macejak,D.G.,
Holecek,J.J. and Mamone,J.A.
TITLE Method and reagent for inhibiting viral replication
JOURNAL Patent: JP 2000342285-A 238 12-DEC-2000;
RIBOZYME PHARMACEUTICALS INC
COMMENT OS Artificial Sequence
PN JP 2000342285-A/238
PD 12-DEC-2000
PF 01-MAY-2000 JP 2000132616
PR 11-MAY-1992 US 07/882713,14-MAY-1992 US 07/882712 PR
14-MAY-1992 US 07/882713,14-MAY-1992 US 07/882714 PR
14-MAY-1992 US 07/882823,14-MAY-1992 US 07/882824 PR
14-MAY-1992 US 07/882886,14-MAY-1992 US 07/882888 PR
14-MAY-1992 US 07/882889,14-MAY-1992 US 07/882921 PR
14-MAY-1992 US 07/882922,14-MAY-1992 US 07/883823 PR
14-MAY-1992 US 07/883849,14-MAY-1992 US 07/884073 PR
14-MAY-1992 US 07/884074,14-MAY-1992 US 07/884333 PR
14-MAY-1992 US 07/884422,14-MAY-1992 US 07/884431 PR
14-MAY-1992 US 07/884436,14-MAY-1992 US 07/884521 PR
31-JUL-1992 US 07/923738,26-AUG-1992 US 07/935854 PR
26-AUG-1992 US 07/936086,18-SEP-1992 US 07/948359 PR
15-OCT-1992 US 07/963332,07-DEC-1992 US 07/987129 PR
07-DEC-1992 US 07/987130,07-DEC-1992 US 07/987133 PI
KENNETH G DRAPER,LEC W DADYKIZ,JAMES A MACSWIGEN, PI DENNIS G
MAYSEJAK,
PI JAMES J HOLESEK,ANTHONY J MAMONE
PC C12N15/09,C12N5/10,C12N7/00,C12N9/22// (C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00, (C12N5/00,C12R1:91)
CC
CC key Location/Qualifiers
FT source 1..16
FT Location/Qualifiers
FT /organism='Artificial Sequence'.
FEATURES
SOURCE 1..16
/organism='synthetic construct'
/mol_type='genomic RNA'
/db_xref='taxon:32630'
BASE COUNT 7 a 1 c 5 g 3 t
ORIGIN

Alignment Scores:
Pred. No.: 158 Length: 16
Score: 20.00 Matches: 4

Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x BD001078 (1-16)

OY 1 LysAspGluLeu 4
 2 AAGATGAACCTT 13

RESULT 4
 LOCUS BD001507 16 bp RNA linear PAT 31-JAN-2002

DEFINITION Method and reagent for inhibiting viral replication.

ACCESSION BD001507
 VERSION BD001507.1 GI:18626066

KEYWORDS JP 2000342286-A/238.
 SOURCE synthetic construct

ORGANISM artificial sequences.
 1 (bases 1 to 16)
 Draper,K.G., Dadykiz,L.W., Macswigen,J.A., Maysejak,D.G.,
 Holesek,J.J. and Mamone,A.J.

TITLE Method and reagent for inhibiting viral replication
 JOURNAL Patent: JP 2000342286-A 238 12-DEC-2000;
 RIBOZYME PHARMACEUTICALS INC

REFERENCE OS Artificial Sequence
 PN JP 2000342286-A/238

PR 11-MAY-1992 US 07/882689,14-MAY-1992 US 07/882712 PR
 PD 12-DEC-2000

14-MAY-1992 US 07/882713,14-MAY-1992 US 07/882824 PR
 14-MAY-1992 US 07/882822,14-MAY-1992 US 07/882824 PR

14-MAY-1992 US 07/882886,14-MAY-1992 US 07/882888 PR
 14-MAY-1992 US 07/882889,14-MAY-1992 US 07/882921 PR

14-MAY-1992 US 07/882922,14-MAY-1992 US 07/883823 PR
 14-MAY-1992 US 07/883849,14-MAY-1992 US 07/884073 PR

14-MAY-1992 US 07/884074,14-MAY-1992 US 07/884313 PR
 14-MAY-1992 US 07/884422,14-MAY-1992 US 07/884431 PR

14-MAY-1992 US 07/884436,14-MAY-1992 US 07/884521 PR
 31-JUL-1992 US 07/921738,26-AUG-1992 US 07/935854 PR

26-AUG-1992 US 07/936086,18-SEP-1992 US 07/948359 PR
 15-OCT-1992 US 07/963322,07-DEC-1992 US 07/987129 PR

07-DEC-1992 US 07/987130,07-DEC-1992 US 07/987133 PR
 KENNETH G DRAPER,LEC W DADYKIZ,JAMES A MACSWIGEN, PI DENNIS G
 WAYSEJAK,

PI JAMES J HOLESEK,ANTHONY J MAMONE
 PC C12N15/09,C12N5/10,C12N7/00//A61K38/43,A61K39/125,A61K39/13,
 PC A61K39/135,
 PC A61K39/145,A61K39/21,A61K39/23,A61K39/245,A61K39/29,A61K46/00,
 PC A61P1/16,
 PC A61P31/14,A61P31/16,A61P31/18,A61P31/22,A61P35/02,C12Q1/68, PC
 (C12N15/09,C12R1:93),C12N15/00,C12N5/00,A61K37/48,C12N15/00, PC
 C12R1:93)

CC Key location/Qualifiers
 FH source 1.16 location/Qualifiers
 FT source 1.16 location/Qualifiers

FEATURES
 source 1.16 location/Qualifiers
 1.16 /organism="synthetic construct"
 /mol_type="genomic RNA"
 /db_xref="taxon:32630"

BASE COUNT 7 a 1 c 5 g 3 t

ORIGIN

Alignment Scores:
 Pred. No.: 158 Length: 16
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 4
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x BD001507 (1-16)

OY 1 LysAspGluLeu 4
 2 AAGATGAACCTT 13

RESULT 5
 AX215210/c 17 bp mRNA linear PAT 07-SEP-2001

LOCUS AX215210 Sequence 652 from Patent WO0159103.

DEFINITION AX215210

ACCESSION AX215210

VERSION AX215210.1 GI:15525253

KEYWORDS

SOURCE synthetic construct

ORGANISM synthetic construct

REFERENCE 1
 Blatt,L., Mcswigen,J. and Chowitra,B.M.

AUTHORS Method and reagent for the modulation and diagnosis of cd20 and

TITLE nogo gene expression

JOURNAL Patent: WO 0159103-A 652 16-AUG-2001;
 Mcswigen,James (US) ; Chowitra, Bharat M. (US)

FEATURES
 source 1.17 location/Qualifiers
 1.17 /organism="synthetic construct"
 /mol_type="mRNA"
 /db_xref="taxon:32630"
 /note="Nucleic Acid"

BASE COUNT 4 a 2 c 4 g 7 t

ORIGIN

Alignment Scores:
 Pred. No.: 168 Length: 17
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0

US-09-696-872-37 (1-4) x AX215210 (1-17)

OY 1 LysAspGluLeu 4
 16 AAGATGAACCTC 5

RESULT 6
 AX215211/c 17 bp mRNA linear PAT 07-SEP-2001

LOCUS AX215211 Sequence 653 from Patent WO0159103.

DEFINITION AX215211

ACCESSION AX215211

VERSION AX215211.1 GI:15525254

KEYWORDS

SOURCE synthetic construct

ORGANISM synthetic construct

REFERENCE 1
 Blatt,L., Mcswigen,J. and Chowitra,B.M.

AUTHORS Method and reagent for the modulation and diagnosis of cd20 and

TITLE nogo gene expression

JOURNAL Patent: WO 0159103-A 653 16-AUG-2001;
 Mcswigen,James (US) ; Chowitra, Bharat M. (US)

FEATURES
 source 1.17 location/Qualifiers
 1.17 /organism="synthetic construct"
 /mol_type="mRNA"
 /db_xref="taxon:32630"
 /note="Nucleic Acid"

BASE COUNT 5 a 2 c 3 g 7 t

ORIGIN

Alignment Scores:
Pred. No.: 168 Length: 17
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AX215211 (1-17)

Qy 1 LysaepgJlueu 4
15 AAAGATGAATC 4

RESULT 7
LOCUS AX215212 17 bp mRNA linear PAT 07-SEP-2001
DEFINITION Sequence 654 from Patent WO0159103.
ACCESSION AX215212
VERSION AX215212.1 GI:15525255
KEYWORDS
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE 1
AUTHORS Blatt, L., McSwiggen, J. and Chowrita, B.M.
TITLE Method and reagent for the modulation and diagnosis of cd20 and nogo gene expression
JOURNAL Patent: WO 0159103-A 654 16-AUG-2001;
RIBOZYME PHARMACEUTICALS, INC. (US); Blatt, Lawrence (US);
McSwiggen, James (US); Chowrita, Bharat M. (US)
LOCATION/Qualifiers
1.17
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/note="Nucleic Acid"

BASE COUNT 5 a 2 c 4 g 6 t
ORIGIN

Alignment Scores:
Pred. No.: 168 Length: 17
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AX215212 (1-17)

Qy 1 LysaepgJlueu 4
12 AAAGATGAATC 1

RESULT 8
LOCUS AX216059 17 bp mRNA linear PAT 07-SEP-2001
DEFINITION Sequence 1501 from Patent WO0159103.
ACCESSION AX216059
VERSION AX216059.1 GI:15526102
KEYWORDS
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE 1
AUTHORS Blatt, L., McSwiggen, J. and Chowrita, B.M.
TITLE Method and reagent for the modulation and diagnosis of cd20 and nogo gene expression
JOURNAL Patent: WO 0159103-A 1501 16-AUG-2001;
RIBOZYME PHARMACEUTICALS, INC. (US); Blatt, Lawrence (US);
McSwiggen, James (US); Chowrita, Bharat M. (US)
LOCATION/Qualifiers

source 1.17
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/note="Nucleic Acid"

BASE COUNT 6 a 2 c 3 g 6 t
ORIGIN

Alignment Scores:
Pred. No.: 168 Length: 17
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AX216059 (1-17)

Qy 1 LysaepgJlueu 4
14 AAAGATGAATC 3

RESULT 9
LOCUS A58382 19 bp DNA linear PAT 05-MAR-1998
DEFINITION Sequence 11 from Patent WO9637607.
ACCESSION A58382
VERSION A58382.1 GI:3714038
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Braun, A. and Kammerer, S.
TITLE NUCLEIC ACID WITH A POLYMORPHIC DOMAIN OF A B2 BRADYKININ RECEPTOR
JOURNAL Patent: WO 9637607-A 11 28-NOV-1996;
COMMENT ROSCHER ADELBERT (DE)
OTHER PUBLICATION DE 19518931 960919.
LOCATION/Qualifiers
1.19
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 4 t
ORIGIN

Alignment Scores:
Pred. No.: 189 Length: 19
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x A58382 (1-19)

Qy 1 LysaepgJlueu 4
8 AAAGATGAATC 19

RESULT 10
LOCUS ARI60690 20 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 16 from patent US 6255105.
ACCESSION ARI60690
VERSION ARI60690.1 GI:16225163
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 20)
AUTHORS Marchetti, A., Buttitta, F., Smith, G.H. and Callahan, R.

TITLE Nucleotide and deduced amino acid sequences of tumor gene Int6
JOURNAL Patent: US 6255105-A 16 03-JUL-2001;
FEATURES Location/Qualifiers
SOURCE 1..20
BASE COUNT 9 a 4 c 4 g 3 t
ORIGIN /organism="unknown"

Alignment Scores:
Pred. No.: 199 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AR160690 (1-20)

QY 1 LysAspGluLeu 4
Db 7 AAGACGAGCTC 18

RESULT 11
AR183981 20 bp DNA linear PAT 20-APR-2002
LOCUS AR183981
DEFINITION Sequence 16 from patent US 6342392.
ACCESSION AR183981
VERSION AR183981.1 GI:20227950
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20)
AUTHORS Marchetti,A., Buttitta,F., Smith,G.H. and Callahan,R.
TITLE Nucleotide and deduced amino acid sequences of tumor gene Int6
JOURNAL Patent: US 6342392-A 16 29-JAN-2002;
FEATURES Location/Qualifiers
SOURCE 1..20
BASE COUNT 9 a 4 c 4 g 3 t
ORIGIN /organism="unknown"

Alignment Scores:
Pred. No.: 199 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AR183981 (1-20)

QY 1 LysAspGluLeu 4
Db 7 AAGACGAGCTC 18

RESULT 12
AR221427/c 20 bp DNA linear PAT 26-SEP-2002
LOCUS AR221427
DEFINITION Sequence 66 from patent US 6426220.
ACCESSION AR221427
VERSION AR221427.1 GI:23328477
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20)
AUTHORS Bennett,C.F. and Cowseert,L.M.
TITLE Antisense modulation of calreticulin expression
JOURNAL Patent: US 6426220-A 66 30-JUL-2002;
FEATURES Location/Qualifiers
SOURCE 1..20
BASE COUNT 199 Length: 20
Pred. No.: 199
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

BASE COUNT 2 a 7 c 6 g 5 t
ORIGIN
Alignment Scores:
Pred. No.: 199 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AR221427 (1-20)

QY 1 LysAspGluLeu 4
Db 13 AAGACGAGCTG 2

RESULT 13
AR221428 20 bp DNA linear PAT 26-SEP-2002
LOCUS AR221428
DEFINITION Sequence 67 from patent US 6426220.
ACCESSION AR221428
VERSION AR221428.1 GI:23328478
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20)
AUTHORS Bennett,C.F. and Cowseert,L.M.
TITLE Antisense modulation of calreticulin expression
JOURNAL Patent: US 6426220-A 67 30-JUL-2002;
FEATURES Location/Qualifiers
SOURCE 1..20
BASE COUNT 2 a 8 c 4 g 6 t
ORIGIN /organism="unknown"

Alignment Scores:
Pred. No.: 199 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AR221428 (1-20)

QY 1 LysAspGluLeu 4
Db 16 AAGACGAGCTG 5

RESULT 14
AR221429 20 bp DNA linear PAT 26-SEP-2002
LOCUS AR221429
DEFINITION Sequence 68 from patent US 6426220.
ACCESSION AR221429
VERSION AR221429.1 GI:23328479
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20)
AUTHORS Bennett,C.F. and Cowseert,L.M.
TITLE Antisense modulation of calreticulin expression
JOURNAL Patent: US 6426220-A 68 30-JUL-2002;
FEATURES Location/Qualifiers
SOURCE 1..20
BASE COUNT 2 a 8 c 3 g 7 t
ORIGIN /organism="unknown"

Alignment Scores:
Pred. No.: 199 Length: 20

Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AR221429 (1-20)

Qy 1 LysAaPGiLeu 4
 |||||
 Db 19 AAGGACGAGCTG 8

RESULT 15

BD138303 20 bp DNA linear PAT 18-SEP-2002
 LOCUS Antisense modulation of human MDM2 expression.

DEFINITION BD138303

ACCESSION BD138303.1 GI:23233248

VERSION JP 2002508944-A/229.

KEYWORDS unclassified

SOURCE unclassified

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 20)

AUTHORS Miraglia,L.J., Nero,P., Graham,M.J., Monia,B.P. and Cowseert,L.M.

TITLE Antisense modulation of human MDM2 expression

JOURNAL Patent: JP 2002508944-A 229 26-MAR-2002;

ISIS PHARMACEUTICALS INC

OS Unidentified

PN JP 2002508944-A/229

PD 26-MAR-2002

PF 26-MAR-1999 JP 2000538025

PR 26-MAR-1998 US 09/048810

PI LOREN J MIRAGLIA, PAMELA NERO, MARK J GRAHAM, BRETT P MONIA, LEX M

COMMENT

PI COMSERT
 PC C12N15/09,A61K48/00,A61P9/10,A61P17/06,A61P35/00,C07H21/04//
 PC C12Q1/68,
 PC C12N15/00
 CC Strandedness: Single;
 CC Topology: Linear;
 CC Antisense modulation of human MDM2 expression FH Key
 CC Location/Qualifiers
 FT source 1..20
 Location/Qualifiers

FEATURES
 source 1..20
 /organism="Unidentified".
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"

BASE COUNT 8 a 3 c 4 g 5 t
 ORIGIN

Alignment Scores:

Pred. No.: 199 Length: 20
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x BD138303 (1-20)

Qy 1 LysAaPGiLeu 4
 |||||
 Db 2 AAGATGAGCTA 13

Search completed: August 26, 2003, 18:30:19
 Job time : 179.416 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_blue_p2n model

Run on: August 26, 2003, 16:04:40 ; Search time 13.6991 Seconds
(without alignments)
788.208 Million cell updates/sec

Title: US-09-696-872-37
Perfect score: 20
Sequence: 1 KDEL 4

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=x1h
-O=/cgnr_1/USPTO.epool/US09696872/runat_26082003_151137_3213/app_query.fasta_1.462
-DB=N-Geneseq_19Jun03 -QPM=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum2 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=PCT -NOM=score -HEAPSIZE=500 -MITLEN=0 -MAYLEN=2000000000
-USER=US09696872.@CGN_1.1.0.@runat_26082003_151137_3213 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQDENTRY -NEG_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
4: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	12	AA235104	DNA encoding endop
2	20	100.0	12	AA235104	Endoplasmic reticu
3	20	100.0	12	AA235104	Anti-human caliciv
4	20	100.0	12	AA235104	Oligonucleotide SE
5	20	100.0	12	AA235104	Oligonucleotide SE
6	20	100.0	12	AA235104	Oligonucleotide SE
7	20	100.0	12	AA235104	Oligonucleotide SE
8	20	100.0	12	AA235104	Oligonucleotide SE
9	20	100.0	12	AA235104	Oligonucleotide SE
10	20	100.0	12	AA235104	Oligonucleotide SE
11	20	100.0	12	AA235104	Oligonucleotide SE
12	20	100.0	12	AA235104	Oligonucleotide SE
13	20	100.0	12	AA235104	Oligonucleotide SE
14	20	100.0	12	AA235104	Oligonucleotide SE
15	20	100.0	12	AA235104	Oligonucleotide SE
16	20	100.0	12	AA235104	Oligonucleotide SE
17	20	100.0	12	AA235104	Oligonucleotide SE
18	20	100.0	12	AA235104	Oligonucleotide SE
19	20	100.0	12	AA235104	Oligonucleotide SE
20	20	100.0	12	AA235104	Oligonucleotide SE
21	20	100.0	12	AA235104	Oligonucleotide SE
22	20	100.0	12	AA235104	Oligonucleotide SE
23	20	100.0	12	AA235104	Oligonucleotide SE
24	20	100.0	12	AA235104	Oligonucleotide SE
25	20	100.0	12	AA235104	Oligonucleotide SE
26	20	100.0	12	AA235104	Oligonucleotide SE
27	20	100.0	12	AA235104	Oligonucleotide SE
28	20	100.0	12	AA235104	Oligonucleotide SE
29	20	100.0	12	AA235104	Oligonucleotide SE
30	20	100.0	12	AA235104	Oligonucleotide SE
31	20	100.0	12	AA235104	Oligonucleotide SE
32	20	100.0	12	AA235104	Oligonucleotide SE
33	20	100.0	12	AA235104	Oligonucleotide SE
34	20	100.0	12	AA235104	Oligonucleotide SE
35	20	100.0	12	AA235104	Oligonucleotide SE
36	20	100.0	12	AA235104	Oligonucleotide SE
37	20	100.0	12	AA235104	Oligonucleotide SE
38	20	100.0	12	AA235104	Oligonucleotide SE
39	20	100.0	12	AA235104	Oligonucleotide SE
40	20	100.0	12	AA235104	Oligonucleotide SE
41	20	100.0	12	AA235104	Oligonucleotide SE
42	20	100.0	12	AA235104	Oligonucleotide SE
43	20	100.0	12	AA235104	Oligonucleotide SE
44	20	100.0	12	AA235104	Oligonucleotide SE
45	20	100.0	12	AA235104	Oligonucleotide SE

ALIGNMENTS

RESULT 1
AA235104
ID AA235104 standard; DNA; 12 BP.
AA235104;
13-MAR-2000 (first entry)
DNA encoding endoplasmic reticulum retention signal.
Endoplasmic reticulum; retention signal; monocotyledonous plant;
Monocot; rice; wheat; transgenic plant; mammalian polypeptide;
antibody; expression cassette; ss.
Synthetic.
MO9966026-A2.

XX 23-DEC-1999.
PD 99WO-US13584.
XX 15-JUN-1999;
PF 99WO-US13584.
XX 15-JUN-1998;
PR 98US-0089322.
XX (INNE-) INNES CENT JOHN.
XX (CHRI/) CHRISTOU P.
XX
PI Christou P, Strager E, Fischer R, Martin-Vaquero C, Schillberg S;
PI Ma JK;
XX WPI; 2000-097739/08.
XX P-PSDB; AAY32395.
XX
XX Production of mammalian polypeptides, especially antibodies, by
PT monocotyledonous plants -
PS Disclosure; Page 5; 76pp; English.
XX
XX This DNA sequence codes for an endoplasmic reticulum (ER) retention
CC signal (see AAY32395). Levels of mammalian polypeptide expression in
CC transgenic monocotyledonous plants can be enhanced by use of this
CC signal. The invention provides rice, wheat and other monocot plants
CC that have been transformed with expression cassettes for production
CC of mammalian polypeptides, such as antibodies. ER retention signals
CC (see AAY32395-96), 5' untranslated regions and leader peptides are
CC employed in various combinations to provide high expression yield.
CC Plant cell or seed containing polypeptides selected from Fv, Fab,
CC F(ab)2, diabody, dimeric scfv, whole antibody and four-chain
CC secretory antibody are claimed. Multi-chain complexes such as
CC four-chain secretory antibodies are produced by expression of
CC component polypeptides from separate vectors, all introduced into
CC the same cell by transformation.
XX
SQ Sequence 12 BP; 5 A; 2 C; 3 G; 2 T; 0 other;
XX
XX Alignment Scores:
Pred. No.: 138 Length: 12
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
XX
US-09-696-872-37 (1-4) x AAZ35104 (1-12)
XX
QY 1 LysAapGluLeu 4
XX |||||
DB 1 AAAGATGAGCTC 12
XX |||||
XX
XX RESULT 2
ID AAF57883 standard; DNA; 12 BP.
XX AAF57883;
XX
XX 20-APR-2001 (first entry)
XX
XX Endoplasmic reticulum retention signal peptide #1 coding sequence.
DE Endoplasmic reticulum; ER; retention signal; transgenic plant;
XX expression cassette; de.
XX
XX Unidentified.
XX
XX WO200105936-A2.
XX
XX 25-JAN-2001.
XX
XX 18-JUL-2000; 2000WO-US19721.
XX
XX

PR 19-JUL-1999; 99US-0144513.
PR 07-JUL-2000; 2000US-0611736.
XX
XX (INNE-) INNES CENT JOHN.
XX
XX Christou P, Kohli A;
XX
XX WPI; 2001-147332/15.
XX P-PSDB; AAB68569.
XX
XX Producing population of transgenic plants involves transforming plant
PT cell with minimal transgene expression cassette by direct DNA transfer
PT techniques and regenerating transgenic plants from transformed cells -
PS Disclosure; Page 10; 42pp; English.
XX
XX The present invention relates to a method for producing a population of
CC transgenic plants, comprising transforming intact plant cells with
CC minimal transgene expression cassettes by direct DNA transfer techniques,
CC and regenerating a population of transgenic plants from transformed plant
CC cells. The present sequence is the coding sequence for an endoplasmic
CC reticulum (ER) retention signal peptide. This sequence can be used in the
CC expression cassette of the present invention.
XX
SQ Sequence 12 BP; 5 A; 2 C; 3 G; 2 T; 0 other;
XX
XX Alignment Scores:
Pred. No.: 138 Length: 12
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
XX
US-09-696-872-37 (1-4) x AAF57883 (1-12)
XX
QY 1 LysAapGluLeu 4
XX |||||
DB 1 AAAGATGAGCTC 12
XX |||||
XX
XX RESULT 3
ID ABT15850 standard; DNA; 12 BP.
XX ABT15850;
XX
XX 28-MAR-2003 (first entry)
XX
XX Anti-human calicivirus monoclonal antibody related DNA SEQ ID No 24.
DE Transgenic plant; anti-human calicivirus monoclonal antibody; gene; de.
XX
XX Unidentified.
XX
XX JP2002253262-A.
XX
XX 10-SEP-2002.
XX
XX 05-MAR-2001; 2001JP-0060462.
XX
XX 05-MAR-2001; 2001JP-0060462.
XX
XX (FRON-) FRONTIER SCI KK.
XX (HOKK-) HOKKAIDO GREEN BIO KENKYUSHO KK.
XX (SUGI/) SUGIMOTO C.
XX (UEDA/) UEDA I.
XX
XX WPI; 2003-150966/15.
XX P-PSDB; ABJ19274.
XX
XX A new transgenic plant cell useful for producing anti-human calicivirus
PT monoclonal antibody -
XX

PS Disclosure, Page 21-22, 36pp, Japanese.
XX
CC The invention relates to a novel method of a transgenic plant, a plant
CC body cell, a plant body, growth of a plant body, an anti-human
CC calicivirus monoclonal antibody, and the detection of a calicivirus. This
CC polynucleotide sequence represents DNA relating to the human calicivirus
CC detection method of the invention.
XX
SQ Sequence 12 BP; 5 A; 1 C; 3 G; 3 T; 0 other;
XX
Alignment Scores:
Pred. No.: 138 Length: 12
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 25 Gaps: 0
US-09-696-872-37 (1-4) x ABT15850 (1-12)
Oy 1 LysAapGiuleu 4
Db 1 AAGATGACTT 12
RESULT 4
ABC27130 standard; DNA; 13 BP.
XX ABC27130;
XX ABC27130;
XX 20-FEB-2002 (first entry)
XX
XX Oligonucleotide SEQ ID NO 27147 for detecting SNP TSC0007386.
DE
XX SNP, single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS Homo sapiens.
XX WO200177384-A2.
XX 18-OCT-2001.
XX 06-APR-2001; 2001WO-IB00713.
XX 07-APR-2000; 2000DE-1019173.
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-657177/75.
XX
XX Set of oligonucleotides, useful for diagnosis and cell typing, is
PT designed to detect single nucleotide polymorphisms and cytosine
PT methylation status -
XX
XX Claim 1; SEQ ID 27147; 29pp + Sequence Listing; German.
XX
XX This invention describes novel oligonucleotide primers or peptide nucleic
CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
CC and cytosine methylation status in chemically pretreated genomic DNA. The
CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a
CC range of diseases including immune system, gastrointestinal, respiratory,
CC central nervous system, cardiovascular and metabolic disorders. The
CC oligomers are also used for detecting cell type differentiation.
CC ABC00010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and
CC AB100010-AB182073 represent the oligomers described in the invention.
CC NOTE: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 13 BP; 6 A; 0 C; 4 G; 3 T; 0 other;
XX
Alignment Scores:
Pred. No.: 151 Length: 13
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-696-872-37 (1-4) x ABC27130 (1-13)
Oy 1 LysAapGiuleu 4
Db 2 AAGATGAGTTA 13
RESULT 5
ABC27131/C
ID ABC27131 standard; DNA; 13 BP.
XX
XX ABC27131;
XX
XX 20-FEB-2002 (first entry)
XX
XX Oligonucleotide SEQ ID NO 27148 for detecting SNP TSC0007386.
DE
XX SNP, single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS Homo sapiens.
XX WO200177384-A2.
XX 18-OCT-2001.
XX 06-APR-2001; 2001WO-IB00713.
XX 07-APR-2000; 2000DE-1019173.
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-657177/75.
XX
XX Set of oligonucleotides, useful for diagnosis and cell typing, is
PT designed to detect single nucleotide polymorphisms and cytosine
PT methylation status -
XX
XX Claim 1; SEQ ID 27148; 29pp + Sequence Listing; German.
XX
XX This invention describes novel oligonucleotide primers or peptide nucleic
CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
CC and cytosine methylation status in chemically pretreated genomic DNA. The
CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a
CC range of diseases including immune system, gastrointestinal, respiratory,
CC central nervous system, cardiovascular and metabolic disorders. The
CC oligomers are also used for detecting cell type differentiation.
CC AB100010-AB182073 represent the oligomers described in the invention.
CC NOTE: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

Alignment Scores:
Pred. No.: 151 Length: 13
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

```
DB: 23 Gaps: 0
US-09-696-872-37 (1-4) x ABC27131 (1-13)
QY 1 LysAspGluLeu 4
DB 12 AAGGACGAGTTA 1

RESULT 6
ABC27132
ID ABC27132 standard; DNA; 13 BP.
AC ABC27132;
XX
XX
XX 20-FEB-2002 (first entry)
DT
DE Oligonucleotide SEQ ID NO 27149 for detecting SNP TSC0007386.
XX
XX SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
XX peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
XX central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
XX Homo sapiens.
XX
XX WO200177384-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-IB00713.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-657177/75.
XX
XX Set of oligonucleotides, useful for diagnosis and cell typing, is
XX designed to detect single nucleotide polymorphisms and cytosine
XX methylation status -
XX
XX Claim 1; SEQ ID 27149; 29pp + Sequence Listing; German.
XX
XX This invention describes novel oligonucleotide primers or peptide nucleic
XX acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
XX and cytosine methylation status in chemically pretreated genomic DNA. The
XX oligonucleotides are used for diagnosis and/or prognosis of cancer and a
XX range of diseases including immune system, gastrointestinal, respiratory,
XX central nervous system, cardiovascular and metabolic disorders. The
XX oligomers are also used for detecting cell type differentiation.
XX
XX AB100010-AB182073 represent the oligomers described in the invention.
XX NOTE: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 13 BP; 6 A; 1 C; 4 G; 2 T; 0 other;

Alignment Scores:
Pred. No.: 151 Length: 13
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-696-872-37 (1-4) x ABC27132 (1-13)
QY 1 LysAspGluLeu 4
DB 2 AAGGACGAGTTA 13
```

```
RESULT 7
ABC27133/C
ID ABC27133 standard; DNA; 13 BP.
XX
XX ABC27133;
XX
XX 20-FEB-2002 (first entry)
DT
DE Oligonucleotide SEQ ID NO 27150 for detecting SNP TSC0007386.
XX
XX SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
XX peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
XX central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
XX Homo sapiens.
XX
XX WO200177384-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-IB00713.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-657177/75.
XX
XX Set of oligonucleotides, useful for diagnosis and cell typing, is
XX designed to detect single nucleotide polymorphisms and cytosine
XX methylation status -
XX
XX Claim 1; SEQ ID 27150; 29pp + Sequence Listing; German.
XX
XX This invention describes novel oligonucleotide primers or peptide nucleic
XX acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
XX and cytosine methylation status in chemically pretreated genomic DNA. The
XX oligonucleotides are used for diagnosis and/or prognosis of cancer and a
XX range of diseases including immune system, gastrointestinal, respiratory,
XX central nervous system, cardiovascular and metabolic disorders. The
XX oligomers are also used for detecting cell type differentiation.
XX
XX AB100010-AB182073 represent the oligomers described in the invention.
XX NOTE: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 13 BP; 2 A; 4 C; 1 G; 6 T; 0 other;

Alignment Scores:
Pred. No.: 151 Length: 13
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0

US-09-696-872-37 (1-4) x ABC27133 (1-13)
QY 1 LysAspGluLeu 4
DB 12 AAGGACGAGTTA 1

RESULT 8
ABF76114
ID ABF76114 standard; DNA; 13 BP.
XX
XX ABF76114;
XX
XX 22-FEB-2002 (first entry)
DT
XX
```


PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-657177/75.
 XX
 PT Set of oligonucleotides, useful for diagnosis and cell typing, is
 PT designed to detect single nucleotide polymorphisms and cytosine
 PT methylation status -
 XX
 PS Claim 1; SEQ ID 176115; 29pp + Sequence Listing; German.
 XX
 CC This invention describes novel oligonucleotide primers or peptide nucleic
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
 CC and cytosine methylation status in chemically pretreated genomic DNA. The
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a
 CC range of diseases including immune system, gastrointestinal, respiratory,
 CC central nervous system, cardiovascular and metabolic disorders. The
 CC oligomers are also used for detecting cell type differentiation.
 CC ABC00010-ABC99989, ABF00010-ABF99989 and
 CC AB100010-AB182073 represent the oligomers described in the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX DB: Sequence 13 BP; 5 A; 1 C; 5 G; 2 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 151 Length: 13
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-696-872-37 (1-4) x ABF76118 (1-13)
 QY 1 LysAaPGLuLeu 4
 DB 1 AAAGACGAGTTG 12
 RESULT 11
 ABF76119/C
 ID ABF76119 standard; DNA; 13 BP.
 AC ABF76119;
 XX
 DT 22-FEB-2002 (first entry)
 XX
 DE Oligonucleotide SEQ ID NO 176116 for detecting SNP TSC0043719.
 XX
 KW SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
 KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
 KW central nervous system; gastrointestinal; respiratory; immune; metabolic.
 XX
 OS Homo sapiens.
 XX
 PN WO200177384-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-IB00713.
 XX
 PR 07-APR-2000; 2000DE-1019173.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-657177/75.
 XX
 PT Set of oligonucleotides, useful for diagnosis and cell typing, is
 PT designed to detect single nucleotide polymorphisms and cytosine

PT methylation status -
 XX
 PS Claim 1; SEQ ID 176116; 29pp + Sequence Listing; German.
 XX
 CC This invention describes novel oligonucleotide primers or peptide nucleic
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
 CC and cytosine methylation status in chemically pretreated genomic DNA. The
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a
 CC range of diseases including immune system, gastrointestinal, respiratory,
 CC central nervous system, cardiovascular and metabolic disorders. The
 CC oligomers are also used for detecting cell type differentiation.
 CC ABC00010-ABC99989, ABF00010-ABF99989 and
 CC AB100010-AB182073 represent the oligomers described in the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX DB: Sequence 13 BP; 2 A; 5 C; 1 G; 5 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 151 Length: 13
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-696-872-37 (1-4) x ABF76119 (1-13)
 QY 1 LysAaPGLuLeu 4
 DB 13 AAAGACGAGTTG 2
 RESULT 12
 ABH29150
 ID ABH29150 standard; DNA; 13 BP.
 AC ABH29150;
 XX
 DT 22-FEB-2002 (first entry)
 XX
 DE Oligonucleotide SEQ ID NO 229127 for detecting SNP TSC0055897.
 XX
 KW SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
 KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
 KW central nervous system; gastrointestinal; respiratory; immune; metabolic.
 XX
 OS Homo sapiens.
 XX
 PN WO200177384-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-IB00713.
 XX
 PR 07-APR-2000; 2000DE-1019173.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-657177/75.
 XX
 PT Set of oligonucleotides, useful for diagnosis and cell typing, is
 PT designed to detect single nucleotide polymorphisms and cytosine
 PT methylation status -
 XX
 PS Claim 1; SEQ ID 229127; 29pp + Sequence Listing; German.
 XX
 CC This invention describes novel oligonucleotide primers or peptide nucleic
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
 CC and cytosine methylation status in chemically pretreated genomic DNA. The
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a

CC range of diseases including immune system, gastrointestinal, respiratory,
CC central nervous system, cardiovascular and metabolic disorders. The
CC oligomers are also used for detecting cell type differentiation.
CC AB000010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and
CC AB100010-AB12073 represent the oligomers described in the invention.
CC NOTE: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 13 BP; 7 A; 0 C; 3 G; 3 T; 0 other;
Alignment Scores:
Pred. No.: 151 Length: 13
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-696-872-37 (1-4) x ABH29150 (1-13)
Oy 1 LysAspGluLeu 4
Db 2 AAAGATGATTG 13
RESULT 13
ABH29151/c
ID ABH29151 standard; DNA; 13 BP.
XX
AC ABH29151;
XX
DT 22-FEB-2002 (first entry)
XX
DE Oligonucleotide SEQ ID NO 229128 for detecting SNP TSC0055897.
XX
XX SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
XX peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
XX central nervous system; gastrointestinal; respiratory; immune; metabolic.
OS Homo sapiens.
XX
XX WO200177384-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-IB00713.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-657177/75.
XX
XX Set of oligonucleotides, useful for diagnosis and cell typing, is
XX designed to detect single nucleotide polymorphisms and cytosine
XX methylation status -
XX
XX Claim 1; SEQ ID 229128; 29pp + Sequence Listing; German.
XX
XX This invention describes novel oligonucleotide primers or peptide nucleic
XX acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
XX and cytosine methylation status in chemically pretreated genomic DNA. The
XX oligonucleotides are used for diagnosis and/or prognosis of cancer and a
XX range of diseases including immune system, gastrointestinal, respiratory,
XX central nervous system, cardiovascular and metabolic disorders. The
XX oligomers are also used for detecting cell type differentiation.
XX AB000010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and
XX AB100010-AB12073 represent the oligomers described in the invention.
XX NOTE: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 13 BP; 3 A; 3 C; 0 G; 7 T; 0 other;
Alignment Scores:
Pred. No.: 151 Length: 13
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
US-09-696-872-37 (1-4) x ABH29151 (1-13)
Oy 1 LysAspGluLeu 4
Db 12 AAAGATGATTG 1
RESULT 14
ABH29152
ID ABH29152 standard; DNA; 13 BP.
XX
AC ABH29152;
XX
DT 22-FEB-2002 (first entry)
XX
DE Oligonucleotide SEQ ID NO 229129 for detecting SNP TSC0055897.
XX
XX SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
XX peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
XX central nervous system; gastrointestinal; respiratory; immune; metabolic.
OS Homo sapiens.
XX
XX WO200177384-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-IB00713.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-657177/75.
XX
XX Set of oligonucleotides, useful for diagnosis and cell typing, is
XX designed to detect single nucleotide polymorphisms and cytosine
XX methylation status -
XX
XX Claim 1; SEQ ID 229129; 29pp + Sequence Listing; German.
XX
XX This invention describes novel oligonucleotide primers or peptide nucleic
XX acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
XX and cytosine methylation status in chemically pretreated genomic DNA. The
XX oligonucleotides are used for diagnosis and/or prognosis of cancer and a
XX range of diseases including immune system, gastrointestinal, respiratory,
XX central nervous system, cardiovascular and metabolic disorders. The
XX oligomers are also used for detecting cell type differentiation.
XX AB000010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and
XX AB100010-AB12073 represent the oligomers described in the invention.
XX NOTE: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 13 BP; 7 A; 1 C; 3 G; 2 T; 0 other;
Alignment Scores:
Pred. No.: 151 Length: 13
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0
 DB: 23 Gaps: 0
 US-09-696-872-37 (1-4) x ABH29152 (1-13)

QY 1 LysApGjuleu 4
 |||||
 Db 2 AAAGACGAATTG 13

RESULT 15

ABH29153/C
 ID ABH29153 standard; DNA, 13 BP.

AC ABH29153;
 XX

DT 22-FEB-2002 (first entry)
 XX

DE Oligonucleotide SEQ ID NO 229130 for detecting SNP TSC0055897.
 XX

XX SNP, single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
 KM peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
 KW central nervous system; gastrointestinal; respiratory; immune; metabolic.
 XX

OS Homo sapiens.
 XX

PN WO200177384-A2.
 XX

PD 18-OCT-2001.
 XX

PF 06-APR-2001; 2001WO-1B00713.
 XX

PR 07-APR-2000; 2000DE-1019173.
 XX

PA (EPIC-) EPIGENOMICS AG.
 XX

PI Olek A, Piepenbrock C, Berlin K;
 XX

DR WPI, 2001-657177/75.
 XX

PT Set of oligonucleotides, useful for diagnosis and cell typing, is
 PT designed to detect single nucleotide polymorphisms and cytosine
 PT methylation status -

PS Claim 1; SEQ ID 229130; 29pp + Sequence Listing; German.
 XX

CC This invention describes novel oligonucleotide primers or peptide nucleic
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
 CC and cytosine methylation status in chemically pretreated genomic DNA. The
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a
 CC range of diseases including immune system, gastrointestinal, respiratory,
 CC central nervous system, cardiovascular and metabolic disorders. The
 CC oligomers are also used for detecting cell type differentiation.

CC ABC00010-ABC99989, ABF0010-ABF99989, ABH0010-ABH99989 and
 CC ABI00010-ABI82073 represent the oligomers described in the invention.

CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 13 BP; 2 A; 3 C; 1 G; 7 T; 0 other;

Alignment Scores:

Pred. No.:	151	Length:	13
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	23	Gaps:	0

US-09-696-872-37 (1-4) x ABH29153 (1-13)

QY 1 LysApGjuleu 4
 |||||
 Db, 12 AAAGACGAATTG 1

Search completed: August 26, 2003, 17:06:38
 Job time : 14.6991 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 16:55:35 ; Search time 3.15044 Seconds

(without alignments)
560.408 Million cell updates/sec

Title: US-09-696-872-37

Sequence: 1 KDEL 4

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 569978 segs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_epool/US09696872/runat_26082003_151138_3255/app_query.fasta_1.462
-DB=Issued_Patents_NA -QEMT=fastap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdt
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORML=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=200000000
-USER=US09696872@cgn2_1.1.76@runat_26082003_151138_3255 -NCPH=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELLOCK=100 -LONGLOG
-DEV TIMEOUT=130 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	12	4	US-08-983-564A-24
2	20	100.0	16	4	US-09-371-772B-7062
3	20	100.0	18	4	US-09-255-911-38
4	20	100.0	20	3	US-09-280-805-229
5	20	100.0	20	3	US-09-280-805-229
6	20	100.0	20	3	US-08-875-847B-16
7	20	100.0	20	4	US-09-378-843-16
8	20	100.0	20	4	US-09-702-327-66
9	20	100.0	20	4	US-09-702-327-67
10	20	100.0	20	4	US-09-702-327-68
11	20	100.0	21	1	US-08-052-997-6
12	20	100.0	21	1	US-08-684-672-6

13	20	100.0	22	3	US-09-191-852-5	Sequence 5, Appli
14	20	100.0	22	4	US-07-974-409C-315	Sequence 315, App
15	20	100.0	22	4	US-07-974-409C-316	Sequence 316, App
16	20	100.0	22	4	US-07-974-409C-317	Sequence 317, App
17	20	100.0	22	4	US-08-817-906-5	Sequence 5, Appli
18	20	100.0	22	5	PCT-US93-00977-315	Sequence 315, App
19	20	100.0	22	5	PCT-US93-00977-316	Sequence 316, App
20	20	100.0	22	5	PCT-US93-00977-317	Sequence 317, App
21	20	100.0	22	5	PCT-US95-13376-5	Sequence 5, Appli
22	20	100.0	24	4	US-09-471-573A-10	Sequence 10, Appli
23	20	100.0	24	4	US-09-471-573A-20	Sequence 20, Appli
24	20	100.0	25	3	US-09-050-159-109	Sequence 109, App
25	20	100.0	26	3	US-09-191-852-6	Sequence 6, Appli
26	20	100.0	26	5	PCT-US95-13376-6	Sequence 6, Appli
27	20	100.0	27	4	US-08-198-603C-18	Sequence 18, Appli
28	20	100.0	30	4	US-09-471-573A-9	Sequence 9, Appli
29	20	100.0	30	4	US-08-882-164D-42	Sequence 42, Appli
30	20	100.0	31	4	US-08-679-645-394	Sequence 394, App
31	20	100.0	32	3	US-09-243-810-4	Sequence 4, Appli
32	20	100.0	32	3	US-09-191-852-7	Sequence 7, Appli
33	20	100.0	38	4	US-08-817-906-7	Sequence 7, Appli
34	20	100.0	38	5	PCT-US95-13376-7	Sequence 7, Appli
35	20	100.0	41	4	US-08-840-713-7	Sequence 7, Appli
36	20	100.0	42	3	US-08-947-965-56	Sequence 56, Appli
37	20	100.0	43	1	US-07-931-473B-256	Sequence 256, App
38	20	100.0	43	1	US-07-714-131C-256	Sequence 256, App
39	20	100.0	43	1	US-08-412-110-256	Sequence 256, App
40	20	100.0	43	1	US-08-409-442A-256	Sequence 256, App
41	20	100.0	43	1	US-08-469-609A-256	Sequence 256, App
42	20	100.0	43	3	US-09-143-190-256	Sequence 256, App
43	20	100.0	43	4	US-09-502-344-256	Sequence 256, App
44	20	100.0	43	4	US-08-840-713-8	Sequence 8, Appli
45	20	100.0				

ALIGNMENTS

RESULT 1
US-08-983-564A-24
; Sequence 24, Application US/08983564A
; Patent No. 6344600
; GENERAL INFORMATION:
; APPLICANT: Meroc, Bertrand
; APPLICANT: Dieryck, Wilfrid
; APPLICANT: Leneze, Philippe
; APPLICANT: Gaden, Michael
; APPLICANT: Pagnier, Renee-Josée
; APPLICANT: Baudino, Sylvie
; APPLICANT: Poyart, Claude
; TITLE OF INVENTION: METHOD FOR PRODUCING HAEMIN PROTEINS USING PLANT CELLS,
; TITLE OF INVENTION: RESULTING PROTEINS AND PRODUCTS CONTAINING SAME
; FILE REFERENCE: 8076.147USWO
; CURRENT APPLICATION NUMBER: US/08/983,564A
; CURRENT FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: PCT/FR96/01123
; PRIOR FILING DATE: 1996-07-17
; PRIOR APPLICATION NUMBER: 95/08615
; PRIOR FILING DATE: 1995-07-17
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 12
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Homo sapiens
US-08-983-564A-24
Alignment Scores: 32 Length: 12
Pred. No.: 20.00 Matches: 4
Score: 100.00% Conservative: 0
Percent Similarity: 100.00%

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 4

US-09-696-872-37 (1-4) x US-08-983-564A-24 (1-12)

QY 1 LysaApGluLeu 4
DB 1 AAGATGAGCTA 12

RESULT 2
US-09-371-772B-7062
Sequence 7062, Application US/09371772B
Patent No. 6566127
GENERAL INFORMATION:
APPLICANT: Ribozyme Pharmaceuticals, Inc.
APPLICANT: Pavco, Pam
APPLICANT: McSwiggen, Jim
APPLICANT: Stinchcomb, Dan
APPLICANT: Escobedo, Jaime
TITLE OF INVENTION: Method and Reagent for the Treatment of Diseases or Conditions Re
FILE REFERENCE: MBH00, 876-J (237/198)
CURRENT APPLICATION NUMBER: US/09/371, 772B
CURRENT FILING DATE: 1995-08-10
PRIOR APPLICATION NUMBER: US 60/005, 974
PRIOR FILING DATE: 1995-10-26
PRIOR APPLICATION NUMBER: US 08/584, 040
PRIOR FILING DATE: 1996-01-08
NUMBER OF SEQ ID NOS: 14225
SOFTWARE: Patent version 3.0
SEQ ID NO 7062
LENGTH: 16
TYPE: RNA
ORGANISM: Homo sapiens
US-09-371-772B-7062

Alignment Scores:
Pred. No.: 44.1 Length: 16
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 4

US-09-696-872-37 (1-4) x US-09-371-772B-7062 (1-16)

QY 1 LysaApGluLeu 4
DB 5 AAGAGCAACUU 16

RESULT 3
US-09-255-911-38
Sequence 38, Application US/09255911
Patent No. 6013522
GENERAL INFORMATION:
APPLICANT: Brett P. Monia
APPLICANT: Lex M. Cowart
TITLE OF INVENTION: ANTISENSE MODULATION OF SMAD1 EXPRESSION
FILE REFERENCE: RTS-0040
CURRENT APPLICATION NUMBER: US/09/255, 911
CURRENT FILING DATE: 1999-02-23
NUMBER OF SEQ ID NOS: 46
SEQ ID NO 38
LENGTH: 18
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Antisense Oligonucleotide
US-09-255-911-38

Alignment Scores:
Pred. No.: 50.3 Length: 18

Score: 20.00 Mismatches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 3

US-09-696-872-37 (1-4) x US-09-255-911-38 (1-18)

QY 1 LysaApGluLeu 4
DB 1 AAGATGAGCTC 12

RESULT 4
US-09-280-805-229
Sequence 229, Application US/09280805
Patent No. 6184212
GENERAL INFORMATION:
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.
APPLICANT: Graham, Brett P. Monia
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDM2
FILE REFERENCE: EXPRESSION
NUMBER OF SEQUENCES: 271
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Jane Massey Licata
STREET: 66 East Main Street
CITY: Marlton
STATE: NJ
COUNTRY: U.S.A.
ZIP: 08053
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PC
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/280, 805
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/048, 810
FILING DATE: March 26, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Licata, Jane Massey
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0346
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-810-1515
TELEFAX: 609-810-1454
INFORMATION FOR SEQ ID NO: 229:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: Yes
US-09-280-805-229

Alignment Scores:
Pred. No.: 56.6 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 3

US-09-696-872-37 (1-4) x US-09-280-805-229 (1-20)

QY 1 LysaApGluLeu 4
DB 2 AAGATGAGCTA 13

RESULT 5
US-09-280-805-230

Sequence 230, Application US/09280805
Patent No. 6184212
GENERAL INFORMATION:
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.
APPLICANT: Graham, Brett P. Monia
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDM2
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 271
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Jane Massey Licata
STREET: 66 East Main Street
CITY: Marlton
STATE: NJ
COUNTRY: U.S.A.
ZIP: 08053
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PC
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/280,805
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/048,810
FILING DATE: March 26, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Licata, Jane Massey
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0346
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-810-1515
TELEFAX: 609-810-1454
INFORMATION FOR SEQ ID NO: 230:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: Yes
US-09-280-805-230

Alignment Scores:
Pred. No.: 56.6 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-696-872-37 (1-4) x US-09-280-805-230 (1-20)

QY 1 LysAspGluLeu 4
Db 6 AAAGTGGCTA 17

RESULT 6
US-08-875-847B-16
Sequence 16, Application US/08875847B
Patent No. 6255105
GENERAL INFORMATION:
APPLICANT: The Government of the United
APPLICANT: States of America as represented by the
APPLICANT: Secretary, Department of Health and Human
APPLICANT: Services; Callahan, Robert; Marchetti,
APPLICANT: Antonio; Buttitta, Fiama; Smith, Gilbert H.
TITLE OF INVENTION: Nucleotide And Deduced
TITLE OF INVENTION: Amino Acid Sequences Of A New Tumor Gene,
TITLE OF INVENTION: Int6, And the Use Of Reagents Derived From
TITLE OF INVENTION: These Sequences In Diagnostic Assays,
TITLE OF INVENTION: Vaccines, Immunotherapy And Gene Therapy
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,847B
FILING DATE: 09-FEB-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/385,998
FILING DATE: 09-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: William S. Feiler
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4179PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-875-847B-16

Alignment Scores:
Pred. No.: 56.6 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-696-872-37 (1-4) x US-08-875-847B-16 (1-20)

QY 1 LysAspGluLeu 4
Db 7 AAAGACGACTC 18

RESULT 7
US-09-378-842-16
Sequence 16, Application US/09378842
Patent No. 6342392
GENERAL INFORMATION:
APPLICANT: The Government of the United
APPLICANT: States of America as represented by the
APPLICANT: Secretary, Department of Health and Human
APPLICANT: Services; Callahan, Robert; Marchetti,
APPLICANT: Antonio; Buttitta, Fiama; Smith, Gilbert H.
TITLE OF INVENTION: Nucleotide And Deduced
TITLE OF INVENTION: Amino Acid Sequences Of A New Tumor Gene,
TITLE OF INVENTION: Int6, And the Use Of Reagents Derived From
TITLE OF INVENTION: These Sequences In Diagnostic Assays,
TITLE OF INVENTION: Vaccines, Immunotherapy And Gene Therapy
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

```
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: FLOPPY DISK
/ COMPUTER: IBM PC COMPATIBLE
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: MS WORD 97
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/378,842
/ FILING DATE:
/ CLASSIFICATION:
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US/08/875,847
/ FILING DATE: 09-FEB-1996
/ APPLICATION NUMBER: 08/385,998
/ FILING DATE: 09-FEB-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: William S. Feller
/ REGISTRATION NUMBER: 26,728
/ REFERENCE/DOCKET NUMBER: 2026-4179PCT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ TELEX: 421792
/ INFORMATION FOR SEQ ID NO: 16:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 20 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
US-09-378-842-16

Alignment Scores:
Pred. No.: 56.6      Length: 20
Score: 20.00         Matches: 4
Percent Similarity: 100.00%  Conservatve: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 4                Gaps: 0

US-09-696-872-37 (1-4) x US-09-378-842-16 (1-20)

QY 1 LysaApGJuleu 4
   |||||
Db 7 AAAGACGAAGCTC 18

RESULT 8
US-09-702-327-66/c
/ Sequence 66, Application US/09702327
/ Patent No. 6426220
/ GENERAL INFORMATION:
/ APPLICANT: C. Frank Bennett
/ TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
/ FILE REFERENCE: RTS-0097
/ CURRENT APPLICATION NUMBER: US/09/702,327
/ CURRENT FILING DATE: 2000-10-30
/ NUMBER OF SEQ ID NOS: 89
/ SEQ ID NO 66
/ LENGTH: 20
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Antisense Oligonucleotide
/
US-09-702-327-66

Alignment Scores:
Pred. No.: 56.6      Length: 20
Score: 20.00         Matches: 4
Percent Similarity: 100.00%  Conservatve: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 4                Gaps: 0

US-09-696-872-37 (1-4) x US-09-702-327-66 (1-20)
```

```
QY 1 LysaApGJuleu 4
   |||||
Db 13 AAAGACGAGCTG 2

RESULT 9
US-09-702-327-67/c
/ Sequence 67, Application US/09702327
/ Patent No. 6426220
/ GENERAL INFORMATION:
/ APPLICANT: C. Frank Bennett
/ TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
/ FILE REFERENCE: RTS-0097
/ CURRENT APPLICATION NUMBER: US/09/702,327
/ CURRENT FILING DATE: 2000-10-30
/ NUMBER OF SEQ ID NOS: 89
/ SEQ ID NO 67
/ LENGTH: 20
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Antisense Oligonucleotide
/
US-09-702-327-67

Alignment Scores:
Pred. No.: 56.6      Length: 20
Score: 20.00         Matches: 4
Percent Similarity: 100.00%  Conservatve: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 4                Gaps: 0

US-09-696-872-37 (1-4) x US-09-702-327-67 (1-20)

QY 1 LysaApGJuleu 4
   |||||
Db 16 AAGACGAGCTG 5

RESULT 10
US-09-702-327-68/c
/ Sequence 68, Application US/09702327
/ Patent No. 6426220
/ GENERAL INFORMATION:
/ APPLICANT: C. Frank Bennett
/ TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
/ FILE REFERENCE: RTS-0097
/ CURRENT APPLICATION NUMBER: US/09/702,327
/ CURRENT FILING DATE: 2000-10-30
/ NUMBER OF SEQ ID NOS: 89
/ SEQ ID NO 68
/ LENGTH: 20
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Antisense Oligonucleotide
/
US-09-702-327-68

Alignment Scores:
Pred. No.: 56.6      Length: 20
Score: 20.00         Matches: 4
Percent Similarity: 100.00%  Conservatve: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 4                Gaps: 0

US-09-696-872-37 (1-4) x US-09-702-327-68 (1-20)

QY 1 LysaApGJuleu 4
   |||||
Db 19 AAGACGAGCTG 8
```

RESULT 11
US-08-052-997-6
; Sequence 6, Application US/08052997
; Patent No. 5556786
; GENERAL INFORMATION:
; APPLICANT: Kere, Juba
; APPLICANT: Schlessinger, David
; APPLICANT: de la Chapelle, Albert
; TITLE OF INVENTION: ANHIDROTIC ECTODERMAL DYSPLASIA GENE
; TITLE OF INVENTION: AND METHOD OF DETECTING SAME
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: POPHAM HAIR SCHNOBRICH & KAUFMAN, LTD.
; STREET: 1225 Eye Street N.W., Suite 1000
; CITY: Washington, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/052,997
; FILING DATE: 27-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Shaughnessy, Brian P.
; REGISTRATION NUMBER: 32,747
; REFERENCE/DOCKET NUMBER: 9594/81-2189
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 289-1200
; TELEFAX: (202) 289-6674
; TELEX: 248516
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: yeast artificial chromosome
; US-08-052-997-6
; Alignment Scores:
; Pred. No.: 59.8 Length: 21
; Score: 20.00 Matches: 4
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 100.00% Indels: 0
; DB: 1 Gaps: 0
US-09-696-872-37 (1-4) x US-08-052-997-6 (1-21)
Qy 1 Lysaapgiuleu 4
Db 8 AAAGATGAGCTG 19
RESULT 12
US-08-684-672-6
; Sequence 6, Application US/08684672
; Patent No. 5700926
; GENERAL INFORMATION:
; APPLICANT: Kere, Juba
; APPLICANT: Schlessinger, David
; APPLICANT: de la Chapelle, Albert
; APPLICANT: SRIVASTAVA, Anand Kumar
; TITLE OF INVENTION: MOLECULAR CLONING OF THE ANHIDROTIC
; TITLE OF INVENTION: ECTODERMAL DYSPLASIA GENE

NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,672
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,997
; FILING DATE: 27-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Shaughnessy, Brian P.
; REGISTRATION NUMBER: 32,747
; REFERENCE/DOCKET NUMBER: 030956-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-684-672-6
; Alignment Scores:
; Pred. No.: 59.8 Length: 21
; Score: 20.00 Matches: 4
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 100.00% Indels: 0
; DB: 1 Gaps: 0
US-09-696-872-37 (1-4) x US-08-684-672-6 (1-21)
Qy 1 Lysaapgiuleu 4
Db 8 AAAGATGAGCTG 19
RESULT 13
US-09-191-852-5
; Sequence 5, Application US/09191852
; Patent No. 6194560
; GENERAL INFORMATION:
; APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,852
; FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13376
FILING DATE: 24-OCT-1995
APPLICATION NUMBER: 08/817,906
FILING DATE: 04-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fox, David L.
REGISTRATION NUMBER: 40,612
REFERENCE/DOCKET NUMBER: P01590US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-651-5151
TELEFAX: 713-651-5246
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-191-852-5

Alignment Scores:
Pred. No.: 63 Length: 22
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-696-872-37 (1-4) x US-09-191-852-5 (1-22)

QY 1 Lysaepgjluleu 4
DB 8 AAAGATGAGCTA 19

RESULT 14
US-07-974-409C-315/C
Sequence 315, Application US/07974409C
Patent No. 6300058
GENERAL INFORMATION:
APPLICANT: Akitaya, Tatsuo
APPLICANT: Mitsuhashi, Masato
APPLICANT: Cooper, Allan
TITLE OF INVENTION: METHOD AND REAGENT
TITLE OF INVENTION: FOR MEASURING MESSENGER RNA
NUMBER OF SEQUENCES: 457
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe, Martens, Olson, and Bear
STREET: 620 Newport Center Dr. Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/974,409C
FILING DATE: 12-NOV-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: HITACHI.006CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
INFORMATION FOR SEQ ID NO: 315:
SEQUENCE CHARACTERISTICS:
LENGTH: 22
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-974-409C-315

Alignment Scores:
Pred. No.: 63 Length: 22
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-696-872-37 (1-4) x US-07-974-409C-315 (1-22)

QY 1 Lysaepgjluleu 4
DB 15 AAAGACGAAGCTC 4

RESULT 15
US-07-974-409C-316/C
Sequence 316, Application US/07974409C
Patent No. 6300058
GENERAL INFORMATION:
APPLICANT: Akitaya, Tatsuo
APPLICANT: Mitsuhashi, Masato
APPLICANT: Cooper, Allan
TITLE OF INVENTION: METHOD AND REAGENT
TITLE OF INVENTION: FOR MEASURING MESSENGER RNA
NUMBER OF SEQUENCES: 457
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe, Martens, Olson, and Bear
STREET: 620 Newport Center Dr. Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/974,409C
FILING DATE: 12-NOV-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: HITACHI.006CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
INFORMATION FOR SEQ ID NO: 316:
SEQUENCE CHARACTERISTICS:
LENGTH: 22
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-974-409C-316

Alignment Scores:
Pred. No.: 63 Length: 22
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-696-872-37 (1-4) x US-07-974-408C-316 (1-22)

OY 1 LysApGluLeu 4

Db 15 AAGAGCGAACTC 4

Search completed: August 26, 2003, 19:22:50
Job time : 4.15044 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 18:30:26 ; Search time 11.5044 Seconds
(without alignments)
781.669 Million cell updates/sec

Title: US-09-696-872-37
Perfect score: 20
Sequence: 1 KDEI 4

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1517243 seqs, 1124081882 residues
Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPFO_spool/US096872/runat_26082003_151139_3337/app_query.fasta_1.462
-DB=Published Applications_NA -OPMT=faetap -SUFFIX=trmpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pcc -THR MAX=100
-THR MIN=0 -ALIGN=15 -MOBE=LOCAL -OUTFMT=plo -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US096872.@CGN_1_1_291@runat_26082003_151139_3337
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEOUTERT -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PC1_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PC1US_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	12	9	US-09-333-527-1 Sequence 1, Appl

2	20	100.0	12	13	US-10-085-853-24	Sequence 24, Appl
3	20	100.0	12	14	US-10-127-427-1	Sequence 1, Appl
4	20	100.0	15	14	US-10-287-919-1514	Sequence 1514, Ap
5	20	100.0	15	14	US-10-287-919-1515	Sequence 1515, Ap
6	20	100.0	15	14	US-10-287-919-2042	Sequence 2042, Ap
7	20	100.0	15	14	US-10-287-919-2043	Sequence 2043, Ap
8	20	100.0	17	11	US-09-780-533A-652	Sequence 652, App
9	20	100.0	17	11	US-09-780-533A-653	Sequence 653, App
10	20	100.0	17	11	US-09-780-533A-654	Sequence 654, App
11	20	100.0	17	11	US-09-780-533A-1501	Sequence 1501, Ap
12	20	100.0	19	10	US-09-969-373-3751	Sequence 3751, Ap
13	20	100.0	20	9	US-09-752-983-429	Sequence 229, App
14	20	100.0	20	9	US-09-752-983-230	Sequence 230, App
15	20	100.0	20	10	US-09-791-406-66	Sequence 66, Appl
16	20	100.0	20	10	US-09-791-406-67	Sequence 67, Appl
17	20	100.0	20	10	US-09-791-406-68	Sequence 68, Appl
18	20	100.0	20	11	US-09-858-152A-16	Sequence 16, Appl
19	20	100.0	22	10	US-09-969-373-3571	Sequence 3571, Ap
20	20	100.0	23	9	US-09-866-778A-10	Sequence 10, Appl
21	20	100.0	23	9	US-09-866-778A-11	Sequence 11, Appl
22	20	100.0	23	14	US-10-002-292A-7	Sequence 7, Appl
23	20	100.0	24	12	US-10-079-167-3	Sequence 3, Appl
24	20	100.0	24	12	US-10-286-628-37	Sequence 37, Appl
25	20	100.0	25	10	US-09-263-959-1077	Sequence 1077, Ap
26	20	100.0	25	14	US-10-215-112-2977	Sequence 2977, Ap
27	20	100.0	25	14	US-10-215-112-12022	Sequence 12022, A
28	20	100.0	25	14	US-10-098-263B-1698	Sequence 1698, Ap
29	20	100.0	25	14	US-10-098-263B-5107	Sequence 5107, Ap
30	20	100.0	25	14	US-10-098-263B-9595	Sequence 9595, Ap
31	20	100.0	25	14	US-10-098-263B-10231	Sequence 10231, A
32	20	100.0	25	14	US-10-098-263B-10867	Sequence 10867, A
33	20	100.0	25	14	US-10-098-263B-17233	Sequence 17233, A
34	20	100.0	25	14	US-10-098-263B-17234	Sequence 17234, A
35	20	100.0	25	14	US-10-098-263B-27417	Sequence 27417, A
36	20	100.0	25	14	US-10-098-263B-31915	Sequence 31915, A
37	20	100.0	25	14	US-10-098-263B-58278	Sequence 58278, A
38	20	100.0	25	14	US-10-098-263B-58369	Sequence 58369, A
39	20	100.0	25	14	US-10-098-263B-60887	Sequence 60887, A
40	20	100.0	25	14	US-10-098-263B-60997	Sequence 60997, A
41	20	100.0	25	14	US-10-098-263B-66666	Sequence 66666, A
42	20	100.0	25	14	US-10-098-263B-70671	Sequence 70671, A
43	20	100.0	25	14	US-10-098-263B-71415	Sequence 71415, A
44	20	100.0	25	14	US-10-098-263B-76429	Sequence 76429, A
45	20	100.0	25	14	US-10-098-263B-96476	Sequence 96476, A

ALIGNMENTS

RESULT 1
US-09-333-527-1
; Sequence 1, Application US/09333527
; Patent No. US20020078472A1
; GENERAL INFORMATION:
; APPLICANT: Paul CHRISTOU; Eva STROGER; Rainer FISCHER; Carmen MARTIN-VAQUERO; Stef
; TITLE OF INVENTION: METHODS AND MEANS FOR EXPRESSION OF MAMMALIAN POLYPEPTIDES
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,527
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/089,322
FILING DATE: June 15, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Mary Anne Schofield
REGISTRATION NUMBER: 36,669
REFERENCE/DOCKET NUMBER: KL/JIC 202.1 - JEL
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-333-527-1

Alignment Scores:
Pred. No.: 190
Score: 20.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 9
Gaps: 0

US-09-696-872-37 (1-4) x US-09-333-527-1 (1-12)

QY 1 LysAspGluLeu 4
|||
1 AAAGATGAGCTC 12

Db

RESULT 2
US-10-085-853-24
Sequence 24, Application US/10085853
GENERAL INFORMATION:
APPLICANT: Merot, Bertrand
APPLICANT: Dieryck, Wilfrid
APPLICANT: Lenee, Philippe
APPLICANT: Marden, Michael
APPLICANT: Gruber, Veronique
APPLICANT: Pagnier, Renee-Josée
APPLICANT: Baudino, Sylvie
APPLICANT: Poyart, Claude
TITLE OF INVENTION: METHOD FOR PRODUCING HAEMIN PROTEINS USING PLANT
TITLE OF INVENTION: CELLS.
FILE REFERENCE: 8076.147USMO
CURRENT APPLICATION NUMBER: US/10/085,853
CURRENT FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 08/983,564
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: PCT/FR96/01123
PRIOR FILING DATE: 1996-07-17
PRIOR APPLICATION NUMBER: 95/08615
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 12
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Homo sapiens
US-10-085-853-24

Alignment Scores:
Pred. No.: 190
Score: 20.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 13
Gaps: 0

US-09-696-872-37 (1-4) x US-10-085-853-24 (1-12)

QY 1 LysAspGluLeu 4
|||
1 AAAGATGAGCTA 12

Db

RESULT 3
US-10-127-427-1
Sequence 1, Application US/10127427
Publication No. US20030051275A1
GENERAL INFORMATION:
APPLICANT: Paul CHRISTOU; Eva STROGER; Rainer FISCHER; Carmen MARTIN-VAQUERO;
Stefan SC
TITLE OF INVENTION: METHODS AND MEANS FOR EXPRESSION OF MAMMALIAN
POLYPEPTIDES
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/127,427
FILING DATE: 23-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,527
FILING DATE: <Unknown>
APPLICATION NUMBER: 60/089,322
FILING DATE: June 15, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Mary Anne Schofield
REGISTRATION NUMBER: 36,669
REFERENCE/DOCKET NUMBER: KL/JIC 202.1 - JEL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
582015
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-127-427-1

Alignment Scores:
Pred. No.: 190
Score: 20.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 14
Gaps: 0

US-09-696-872-37 (1-4) x US-10-127-427-1 (1-12)

QY 1 LysAspGluLeu 4
|||
1 AAAGATGAGCTC 12

Db

RESULT 4
US-10-287-919-1514/C
Sequence 1514, Application US/10287919
Publication No. US20030085830A1

```
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT APPLICATION NUMBER: US/10/287,919
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 1514
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (809601)...(809615)
; OTHER INFORMATION: Chromosome = 1 Strand = negative ConnectorObjectNumber = 1910
US-10-287-919-1514

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-1514 (1-15)

QY 1 LysAspGluLeu 4
Db 12 AAGGATGAACCTA 1

RESULT 5
US-10-287-919-1515
; Sequence 1515, Application US/10287919
; Publication No. US20030085830A1
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT APPLICATION NUMBER: US/10/287,919
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 1515
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (809601)...(809615)
; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectorObjectNumber = 1905
US-10-287-919-1515

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-1515 (1-15)

QY 1 LysAspGluLeu 4
Db 4 AAGGATGAACCTA 15

RESULT 6
US-10-287-919-2042
; Sequence 2042, Application US/10287919
; Publication No. US20030085830A1
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
```

```
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT APPLICATION NUMBER: US/10/287,919
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 2042
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (1242502)...(1242516)
; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectorObjectNumber = 2612
US-10-287-919-2042

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-2042 (1-15)

QY 1 LysAspGluLeu 4
Db 4 AAGGATGAACCTA 15

RESULT 7
US-10-287-919-2043/c
; Sequence 2043, Application US/10287919
; Publication No. US20030085830A1
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT APPLICATION NUMBER: US/10/287,919
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 2043
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (1242502)...(1242516)
; OTHER INFORMATION: Chromosome = 1 Strand = negative ConnectorObjectNumber = 2611
US-10-287-919-2043

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-2043 (1-15)

QY 1 LysAspGluLeu 4
Db 12 AAGGATGAACCTA 1

RESULT 8
US-09-780-533A-652/c
; Sequence 652, Application US/09780533A
; Publication No. US20030060611A1
; GENERAL INFORMATION:
; APPLICANT: Ribozyme Pharmaceuticals, Inc.
; APPLICANT: Blatt, Larry
; APPLICANT: McSwiggen, Jim
; APPLICANT: Chowrira, Bharat
; APPLICANT: Haeblerli, Pete
```


GENERAL INFORMATION:
APPLICANT: Effertz, Roger J.
APPLICANT: Hauge, Brian M.
TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
FILE REFERENCE: 38-10(52679)A
CURRENT APPLICATION NUMBER: US/09/969,373
PRIOR FILING DATE: 2001-10-02
PRIOR APPLICATION NUMBER: US 09/754,853
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 09/760,427
PRIOR FILING DATE: 2001-01-13
PRIOR APPLICATION NUMBER: US 09/855,768
NUMBER OF SEQ ID NOS: 4593
SEQ ID NO 3751
LENGTH: 19
TYPE: DNA
ORGANISM: Glycine max
US-09-969-373-3751

Alignment Scores:
Pred. No.: 301 Length: 19
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-696-872-37 (1-4) x US-09-969-373-3751 (1-19)

OY 1 LysApGluLeu 4
Db 5 AAGGCGAGATTG 16

RESULT 13
US-09-752-983-229
Sequence 229, Application US/09752983
Patent No. US20010016575A1
GENERAL INFORMATION:
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.
APPLICANT: Graham, Brett P. Montia
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDN2
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 271
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Jane Massey Licata
STREET: 66 East Main Street
CITY: Marilton
STATE: NJ
COUNTRY: U.S.A.
ZIP: 08053
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PC
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/752,983
FILING DATE: 02-Jan-2001
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/280,805
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Licata, Jane Massey
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0346
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-810-1454
TELEFAX: 609-810-1454
INFORMATION FOR SEQ ID NO: 229:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs

TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: Yes
US-09-752-983-229

Alignment Scores:
Pred. No.: 317 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x US-09-752-983-229 (1-20)

OY 1 LysApGluLeu 4
Db 2 AAGGATGAGCTA 13

RESULT 14
US-09-752-983-230
Sequence 230, Application US/09752983
Patent No. US20010016575A1
GENERAL INFORMATION:
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.
APPLICANT: Graham, Brett P. Montia
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDN2
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 271
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Jane Massey Licata
STREET: 66 East Main Street
CITY: Marilton
STATE: NJ
COUNTRY: U.S.A.
ZIP: 08053
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PC
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/752,983
FILING DATE: 02-Jan-2001
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/280,805
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Licata, Jane Massey
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: ISPH-0346
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-810-1454
TELEFAX: 609-810-1454
INFORMATION FOR SEQ ID NO: 230:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: Yes
US-09-752-983-230

Alignment Scores:
Pred. No.: 317 Length: 20
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x US-09-752-983-230 (1-20)

Qy 1 LysAspGluLeu 4
|||||
Db 6 AAGATGAGCTA 17

RESULT 15

US-09-791-406-66/c
; Sequence 66, Application US/09791406
; Patent No. US20020147165A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Robert Rothlein
; APPLICANT: Takashi Kei Kishimoto
; APPLICANT: Lex M. Cowseart
; TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
; FILE REFERENCE: RTS-0097
; CURRENT APPLICATION NUMBER: US/09/791,406
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 66
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Antisense Oligonucleotide
US-09-791-406-66

Alignment Scores:

Pred. No.:	317	Length:	20
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-696-872-37 (1-4) x US-09-791-406-66 (1-20)

Qy 1 LysAspGluLeu 4
|||||
Db 13 AAGAGCGAGCTG 2

Search completed: August 26, 2003, 21:02:50
Job time : 12.504 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 26, 2003, 16:41:45 ; Search time 107.894 Seconds
(without alignments)
901.052 Million cell updates/sec

Title: US-09-696-872-37
Perfect score: 20
Sequence: 1 KDEL 4

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2281392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_epool/US09696872/runat_26082003_151137_3236/app_query.fasta_1.462
-DB=EST -QFMT=fastap -SUFFIX=ret -MINMATCH=0.1 -DOORCL=0 -DOOREXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09696872_QCGN_1.1_3586 @runat_26082003_151137_3236 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DISPLOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fut:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	20	100.0	25	9	AU254198	AU254198 AU254198
2	20	100.0	32	14	D19573	D19573 MUSGS00978
3	20	100.0	33	9	AU254805	AU254805 AU254805
4	20	100.0	38	9	AL801834	AL801834 AL801834
5	20	100.0	49	9	AI130296	AI130296 SMOV33CAN
6	20	100.0	51	9	AI208498	AI208498 qg55h03.x
7	20	100.0	51	29	B2766501	B2766501 SALK_1374
8	20	100.0	52	9	AV555282	AV555282 AV555282
9	20	100.0	52	28	BH620522	BH620522 1007066F1
10	20	100.0	53	9	AA837889	AA837889 oea4c01.s
11	20	100.0	56	9	AU259239	AU259239 AU259239
12	20	100.0	59	29	B2769592	B2769592 SALK_1424
13	20	100.0	60	12	BJ080266	BJ080266 BJ080266
14	20	100.0	60	28	BH011469	BH011469 BG02177-5
15	20	100.0	61	9	AV533632	AV533632 AV533632
16	20	100.0	61	29	AL942430	AL942430 Arabidops
17	20	100.0	62	14	TS7059	TS7059 yc20a01.s1
18	20	100.0	63	28	BH857575	BH857575 SALK_0170
19	20	100.0	64	9	AM420167	AM420167 Fj88b03.y
20	20	100.0	64	12	B1142411	B1142411 SMOV33CAM
21	20	100.0	64	12	B1142469	B1142469 SMOV33CAM
22	20	100.0	65	14	CA840831	CA840831 pf60a03.y
23	20	100.0	66	13	BO519722	BO519722 r439h08.y
24	20	100.0	67	9	AA519033	AA519033 v126d01.r
25	20	100.0	69	28	BH856618	BH856618 SALK_0794
26	20	100.0	70	28	BH856626	BH856626 SALK_0794
27	20	100.0	70	28	BH856627	BH856627 SALK_0794
28	20	100.0	70	28	BH856633	BH856633 SALK_0794
29	20	100.0	70	28	BH856634	BH856634 SALK_0794
30	20	100.0	70	28	BH856640	BH856640 SALK_0793
31	20	100.0	71	9	AV533656	AV533656 AV533656
32	20	100.0	71	13	BU672589	BU672589 haeg002xj
33	20	100.0	72	9	AW542189	AW542189 C0154F09-3
34	20	100.0	73	10	BE420479	BE420479 SMOV2CAS
35	20	100.0	73	14	CB832348	CB832348 SMOV33CAN
36	20	100.0	73	14	CB832443	CB832443 SMOV33CAN
37	20	100.0	73	14	CB832487	CB832487 SMOV33CAN
38	20	100.0	73	14	CB832565	CB832565 SMOV33CAN
39	20	100.0	74	13	BO613177	BO613177 r402d06.y
40	20	100.0	77	9	AI952392	AI952392 wx73b08.x
41	20	100.0	78	12	B1065092	B1065092 p9f1n.pk0
42	20	100.0	78	12	B1468738	B1468738 sa102026.f
43	20	100.0	78	28	AZ310739	AZ310739 IM0025P15
44	20	100.0	79	10	BE316173	BE316173 NF033C07L
45	20	100.0	79	28	AZ614196	AZ614196 IM0442124

ALIGNMENTS

RESULT 1
LOCUS AU254198 25 bp mRNA linear EST 25-APR-2002
DEFINITION AU254198 3'-directed mouse cDNA library Mus musculus cDNA clone
ACCESSION BED0000932 3', mRNA sequence.
VERSION AU254198
KEYWORDS AU254198.1 GI:20315754
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 25)

AUTHORS Kato, K. and Matoba, R.
TITLE Generation of expressed sequence tags from mouse brain
JOURNAL Unpublished
COMMENT Contact: Kikuya Kato
Graduate School of Biological Sciences
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0101, Japan
Tel: 81-743-72-5581
Fax: 81-743-72-5589
Email: kkatobs.aist-nara.ac.jp,
URL: <http://love2.aist-nara.ac.jp/BED/index.html>.

FEATURES
source
1.25
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="BED0000932"
/tissue_type="brain"
/clone_lib="3'-directed mouse cDNA library"
Location/Qualifiers

BASE COUNT 13 a 3 c 4 g 5 t
ORIGIN

Alignment Scores:

Pred. No.: 1.53e+03 Length: 25
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AU254198 (1-25)

Qy 1 LysAepGjuleu 4
|||||
Db 10 AAAGATGAATTG 21

RESULT 2
D19573 32 bp mRNA linear EST 12-DEC-1995
LOCUS MUSGS00978 Mouse 3'-directed Mus musculus domesticus cDNA clone
DEFINITION mb0732 3', mRNA sequence.
D19573
ACCESSION D19573.1 GI:1089429
VERSION
KEYWORDS EST.
SOURCE Mus musculus domesticus (western European house mouse)
ORGANISM Mus musculus domesticus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 32)
AUTHORS Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara, K.
TITLE Analysis of gene expression in mouse embryogenesis by 3'-directed
CDNA sequencing
JOURNAL Unpublished
COMMENT Contact: Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara
, K.
Institute for Cellular and Molecular Biology
Osaka University
3-1 Yamada-oka, Suita, Osaka 565, Japan.
Location/Qualifiers

FEATURES
source
1.32
/organism="Mus musculus domesticus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10092"
/clone="mb0732"
/tissue_type="decidual tissue (day 6.5-8.5 of gestation)"
/clone_lib="mouse 3'-directed"
BASE COUNT 14 a 4 c 5 g 9 t
ORIGIN

Alignment Scores:
Pred. No.: 2.14e+03 Length: 32
Score: 20.00 Matches: 4

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x D19573 (1-32)

Qy 1 LysAepGjuleu 4
|||||
Db 10 AAAGATGAATTG 21

RESULT 3
AU254805 33 bp mRNA linear EST 25-APR-2002
LOCUS AU254805 3'-directed mouse cDNA library Mus musculus cDNA clone
DEFINITION BED000418 3', mRNA sequence.
AU254805
ACCESSION AU254805.1 GI:20316936
VERSION
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 33)
AUTHORS Kato, K. and Matoba, R.
TITLE Generation of expressed sequence tags from mouse brain
JOURNAL Unpublished
COMMENT Contact: Kikuya Kato
Graduate School of Biological Sciences
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0101, Japan
Tel: 81-743-72-5581
Fax: 81-743-72-5589
Email: kkatobs.aist-nara.ac.jp,
URL: <http://love2.aist-nara.ac.jp/BED/index.html>.

FEATURES
source
1.33
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="BED000418"
/tissue_type="brain"
/clone_lib="3'-directed mouse cDNA library"
BASE COUNT 17 a 3 c 4 g 9 t
ORIGIN

Alignment Scores:
Pred. No.: 2.23e+03 Length: 33
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AU254805 (1-33)

Qy 1 LysAepGjuleu 4
|||||
Db 10 AAAGATGAATTG 21

RESULT 4
AL801834 38 bp mRNA linear EST 27-JUN-2002
LOCUS AL801834 XGC-neurula Silurana tropicalis cDNA clone TNeu097103 5',
DEFINITION mRNA sequence.
AL801834
ACCESSION AL801834.1 GI:21588202
VERSION
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
Xenopodinae; Silurana.

REFERENCE 1 (bases 1 to 38)
 AUTHORS Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.
 TITLE Sanger Xenopus tropicalis EST project 2002
 JOURNAL Unpublished
 COMMENT Contact: Taylor R
 Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE_ID: TNeu097103.p1csp6
 Sequencing primer: F1CSP6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 Location/Qualifiers
 1..38
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TNeu097103"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /clone_lib="XGC-neurula"
 /note="Vector: pCS107; Site_1: EcoRI, Site_2: NotI; cDNA was oligo dt primed from 5' end of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."
 BASE COUNT 11 a 9 c 9 g 9 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2.7e+03 Length: 38
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AL601834 (1-38)

QY 1 LysApGtluen 4
 |||||
 16 AAGAGCGAATTA 27

RESULT 5
 A1130296 49 bp mRNA linear EST 26-MAR-1999
 LOCUS SMOVL3CAN08G12 Onchocerca volvulus infective larva cDNA
 DEFINITION (SMM94ML-OVL3) Onchocerca volvulus cDNA clone onch276 5' similar to TR:O01657 O01657 FRUCTOSE-BISPHOSPHATE ALDOLASE ;, mRNA sequence.
 A1130296
 ACCESSION A1130296.1 GI:3600314
 VERSION
 KEYWORDS EST.
 SOURCE Onchocerca volvulus
 ORGANISM Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.
 1 (bases 1 to 49)
 Williams, S.A., Lizotte-Maniewski, M., Laney, S., Wenhong, L., Hillier, L., Allen, M., Bowles, L., Geisel, S., Jost, S., Kucaba, T., Martin, J., Steptoe, M., Theising, B., White, Y., Wylie, T., Chappell, J., Pearson, B., Gibbons, M., Harvey, N., Page, D., Chamberlain, A., Morales, R., Schurk, R., Rittler, E., Kohn, S., Underwood, K. and Maira, M.
 Molecular Parasitology OVL3
 Unpublished
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 The library was constructed by Wenhong Lu. The library is available

from Dr. S.A. Williams, email genome@smith.edu when requesting this
 clone from Dr. Williams, please reference the Williams lab clone id
 - SMOVL3CAN08G12
 Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers
 1..49
 /organism="Onchocerca volvulus"
 /mol_type="mRNA"
 /strain="Sierra Leone"
 /db_xref="taxon:6282"
 /clone="onch276"
 /lab_host="XLI-Blue MRF"
 /clone_lib="Onchocerca volvulus infective larva cDNA (SMM94ML-OVL3)"
 /note="Vector: lambda unizap XR; Site_1: EcoR I; Site_2: Xho I; Cutaneous filarial nematode parasite of humans. mRNA was prepared from third stage infective larvae of Onchocerca volvulus isolated from mosquitoes 10 days after infection and converted to double stranded cDNA using reverse transcriptase and oligo(dt) followed by RNase H and DNase I. The library had 1.8 x 10E5 independent recombinants and average insert size was 900 base pairs. The library was constructed by Wenhong Lu. The library is available from Dr. S.A. Williams, email genome@smith.edu."
 BASE COUNT 15 a 9 c 11 g 14 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.8e+03 Length: 49
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x A1130296 (1-49)

QY 1 LysApGtluen 4
 |||||
 26 AAAGATGATTTG 37

RESULT 6
 A1208498 51 bp mRNA linear EST 30-NOV-1998
 LOCUS qg5h03.x1 Soares testis NHT Homo sapiens cDNA IMAGE:1839125
 DEFINITION qg5h03.x1 Soares testis NHT Homo sapiens cDNA IMAGE:1839125 3' similar to gb:U73628 KAPPA CASEIN PRECURSOR (HUMAN);, mRNA sequence.
 A1208498
 ACCESSION A1208498.1 GI:3770440
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 51)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgaf.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.
 cDNA library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bdip/image/image.html
 Trace considered overall poor quality

```

Insert Length: 761      Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
      Location/Qualifiers
          1. .51
FEATURES
source
```

BASE COUNT
ORIGIN

21 a 10 c 7 g 13 t

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1839125"
/sex="male"
/lab_host="DH10B"
/clone_id="Soares testis NHT"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech laboratories
Inc. and primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGGCGCCGCCCAATTATTTTATTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalisation to Cots and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

Alignment Scores:	
Pred. No.:	4.01e+03
Score:	20.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	
US-09-696-872-37 (1-4)	×
AI208498 (1-51)	
Length:	51
Matches:	4
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

QY 1 LysAspGluLeu 4
|||
Db 8 AAAGATGAACTA 19

LOCUS	DEFINITION	51 bp	DNA	1 linear	GSS 13-MAR-2003
BZ766501/c	BZ766501				
	SALK_137492.43.00.x Arabidopsis thaliana genomic clone SALK_137492.43.00.x, genomic survey sequence.				

ACCESSION	BZ/66501	
VERSION	BZ766501.1	GI:28939054
KEYWORDS	GSS,	
SOURCE	Arabidopsis thaliana (thale cress)	
ORGANISM	Arabidopsis thaliana	

Eurosid II, Brassicales, Brassicaceae, Arabidopsis.

1045851, 1045175
 REFERENCES
 AUTHORS
 Alonso, J. M., Leisner, T. J., Barajas, P., Chen, H.,
 C. Jeske, A., Kaines, M., Kim, C. J., Parker, H.,
 Zimmerman, J., and Ecker, J. R., Prednis, L., Shinn, P.

TITLE	A sequence-indexed library of insertion mutations in the Arabidopsis Genome
JOURNAL	Unpublished
COMMENT	Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (Signal)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752

Fax: 650 536 6379
 Email: ecker@saik.edu
 This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within 300 bases of the 5' end of
 At1g20200 and 300 bases of the 5' end of At1g20210.
 Class: TDNA tagged.

FEATURES	Location/Qualifiers
source	1. .51

```

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SAUK_137492.43.00.x"
/clone_1b="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/cdna\_protocols.html"

```

Pred. No.:	4.00e+03	Length:	51
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	29	Gaps:	0

US-09-696-872-37 (1-4) x BZ766501 (1-51)

QY 1 LYSASPGLULEU 4
|||
Db 16 AAGGACGAGCTG 5

RESULT 8					
AV555282/c					
LOCUS					
DEFINITION	AV555282	52 bp	mRNA	linear	EST 06-SEP-2000
	AV555282		Arabidopsis thaliana green siliques	Columbia	Arabidopsis
	thaliana cDNA clone SC010f02f 3'		mRNA sequence.		

ACCESSION	AV555282	GI:8726697
VERSION	AV555282.1	
KEYWORDS	EST.	
SOURCE	Arabidopsis thaliana (thale cress)	
ORGANISM	Arabidopsis thaliana	

REFERENCE
AUTHORS
TITLE

Aasamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
1 (bases 1 to 52)
A large scale analysis of cDNA in *Arabidopsis thaliana*: Generation
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eustoids II; Brassicales; Brassicaceae; Arabidopsis.

of 12,028 non-redundant expressed size-selected cDNA libraries DNA Res. 7, 175-180 (2000)

MEDLINE	20363093
PUBMED	10907847
COMMENT	Contact: Erika Asamizu

FEATURES
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1533-3, Kisarazu, Chiba 292-0812, Japan
Email: asami@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>
Location/Qualifiers

Source

```

/organism="Arabidopsis thaliana"
/mol_type="rRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ010FC2P"
/tissue_type="green siliques"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/note="Vector: pBluescriptII SK-, Site_1: EcoRI, Site_2:
XhoI"

```

BASE COUNT	21 a	7 c	4 g	20 f
ORIGIN				

Alignment Scores:

Pred. No.:	4.11e+03	Length:	52
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AV555282 (1-52)

QY 1 LysApGluLeu 4
DB 27 AAAGATGACTTA 16

RESULT 9
LOCUS BH620522 52 bp DNA linear GSS 30-JAN-2002
DEFINITION 1007066F12.2EL x1 1007 - Rescuemu Grid H zea mays genomic, genomic survey sequence.

ACCESSION BH620522.1 GI:18432496
VERSION BH620522.1
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
AUTHORS 1 (bases 1 to 52)
TITLE Walbot, V.

JOURNAL Maize genomic sequences found using engineered Rescuemu transposon
COMMENT Unpublished
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu

FEATURES
source Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1007066 column: 35
Class: transposon-tagged.
Location/Qualifiers

1..52
/organism="Zea mays"
/mol_type="Genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1007 - Rescuemu Grid H"
/note="Organ: leaf; Vector: Rescuemu (engineered from Bluescript backbone), Site 1: BamHI; Site 2: BglII; Rescuemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescuemu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'Rescuemu'. Grid H was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."
BASE COUNT 10 a 18 c 9 g 15 t
ORIGIN

Alignment Scores:

Pred. No.: 4.11e+03 Length: 52
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 28 Gaps: 0

US-09-696-872-37 (1-4) x BH620522 (1-52)

QY 1 LysApGluLeu 4
DB 32 AAAGATGAGCTG 21

RESULT 10
LOCUS AA837889 53 bp mRNA linear EST 18-MAR-1998
DEFINITION o040c01.61 NCI CGAP Pr25 Homo sapiens CDNA clone IMAGE:1411008 similar to gb:M73628 KAPPA CASEIN PRECURSOR (HUMAN);, mRNA sequence.

ACCESSION AA837889
VERSION AA837889.1 GI:2913546
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 53)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright, Ph.D.

FEATURES
source CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, at: www-bio.lnl.gov/bbrp/image/image.html
Insert length: 857 Std Error: 0.00
Seq primer: -40m13 fwd, ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers

1..53
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1411008"
/tissue_type="epithelium (cell line)"
/lab_host="SOLR (Kanamycin resistant)"
/clone_lib="NCI CGAP Pr25"
/note="Organ: prostate; Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Normal prostate epithelial cell line (HPV immortalized). 5' adaptor sequence: 5' GAATTCGCGACGAG 3' 3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."
BASE COUNT 21 a 12 c 8 g 11 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 4.22e+03 Length: 53
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AA837889 (1-53)

QY 1 LysApGluLeu 4
DB 8 AAAGATGACTTA 19

RESULT 11
LOCUS AU259239 56 bp mRNA linear EST 25-APR-2002
DEFINITION AU259239 3'-directed mouse cDNA library Mus musculus cDNA clone BED0014799 3', mRNA sequence.
ACCESSION AU259239

VERSION AU259239.1 GI:20325573
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 56)
 Kato, K. and Matsuda, R.
 Generation of expressed sequence tags from mouse brain
 Unpublished
 TITLE Contact: Kikuya Kato
 Graduate School of Biological Sciences
 Nara Institute of Science and Technology
 8916-5 Takeyama, Ikoma, Nara 630-0101, Japan
 Tel: 81-743-72-5581
 Fax: 81-743-72-5589
 Email: kkatoc@bs.ist-nara.ac.jp,
 URL: <http://love2.aist-nara.ac.jp/BED/index.html>.
 FEATURES
 source
 1..56
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="BED0014799"
 /tissue_type="brain"
 /clone_lib="3"-directed mouse cDNA library"
 BASE COUNT 24 a 10 c 7 g 14 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.54e+03 Length: 56
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-696-872-37 (1-4) x AU259239 (1-56)
 QY 1 LysAAGGtLLeu 4
 |||||
 39 AAGATGAATG 28
 RESULT 12 59 bp DNA linear GSS 13-MAR-2003
 BZ769592
 LOCUS SALK_142419.31.10.n Arabidopsis thaliana TDNA insertion line
 DEFINITION Arabidopsis thaliana genomic clone SALK_142419.31.10.n, genomic survey sequence.
 ACCESSION BZ769592
 VERSION BZ769592.1 GI:28943276
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 59)
 Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.U., Parker, H., Prednis, L., Shim, P., Zimmermann, U., and Becker, J.R.
 A Sequence-Indexed library of Insertion Mutations in the Arabidopsis Genome
 Unpublished
 TITLE Contact: Joseph R. Becker
 Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of At1g03370.

FEATURES
 source
 1..59
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_142419.31.10.n"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"
 BASE COUNT 17 a 9 c 14 g 19 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.87e+03 Length: 59
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 29 Gaps: 0
 US-09-696-872-37 (1-4) x BZ769592 (1-59)
 QY 1 LysAAGGtLLeu 4
 |||||
 37 AAGATGAATG 48
 RESULT 13 60 bp mRNA linear EST 12-DEC-2001
 BJ080266
 LOCUS BJ080266
 DEFINITION BJ080266 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL075a15 3', mRNA sequence.
 ACCESSION BJ080266
 VERSION BJ080266.1 GI:17573878
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus.
 1 (bases 1 to 60)
 Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.
 Expressed genes in X. laevis embryo
 Unpublished
 TITLE Contact: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tehin@genes.nig.ac.jp.
 FEATURES
 source
 1..60
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XL075a15"
 /tissue_type="whole embryo"
 /dev_stage="stage 25"
 /clone_lib="NIBB Mochii normalized Xenopus tailbud library"
 BASE COUNT 21 a 9 c 14 g 13 t 3 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.99e+03 Length: 60
 Score: 20.00 Matches: 4

Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 12 Gaps: 0

US-09-696-872-37 (1-4) x BU080266 (1-60)

OY 1 LysApGluLeu 4
 DB 28 AAAGATGAGCTG 39

RESULT 14
 LOCUS BH011469
 DEFINITION BG02177-5prine Drosophila melanogaster P[Gtl] P element insertion
 lines Drosophila melanogaster genomic Sequence recovered from 5'
 end of P element, genomic survey sequence.

ACCESSION BH011469 60 bp DNA linear GSS 11-MAY-2001
 VERSION BH011469
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 60)
 Leyva, R., Hopkins, R., Liao, G., Morzen, N., Tsang, G., He, Y., Karpen
 'G., Bellin, H., Rubin, G. and Spradling, A.
 TITLE The Berkeley Drosophila Genome Project Gene Disruption Project
 JOURNAL Unpublished
 COMMENT Contact: Gerald Rubin
 Berkeley Drosophila Genome Project
 University of California, Berkeley
 LSA Building, Berkeley, CA 94720-3200, USA
 Fax: 5106439947
 Email: gerry@fruitfly.berkeley.edu
 Sequence recovery method was inverse PCR.
 Sequence orientation is forward strand relative to 5' end of P
 element

The P element insertion position is base 53 in the 60 bases. This
 insertion position refers to the first base of the 8 base target
 recognition sequence.
 Class: transposon-tagged.

FEATURES
 source Location/Qualifiers

1..60
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone_lib="Drosophila melanogaster P[Gtl] P element
 insertion lines"
 /note="inverse PCR was performed on Drosophila
 melanogaster strains each of which contains one or more
 P[Gtl] P-element transposon insertion. The resultant
 fragment for each strain was directly sequenced to
 determine the genomic sequence at the site of insertion.
 Details of the protocols used can be found at
 http://www.fruitfly.org/about/methods/inverse.pcr.html."

BASE COUNT 20 a 19 c 16 g 5 t
 ORIGIN

Alignment Scores:
 Pred. No.: 4.99e+03 Length: 60
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 28 Gaps: 0

US-09-696-872-37 (1-4) x BH011469 (1-60)

OY 1 LysApGluLeu 4
 DB 35 AAAGCGAATC 46

RESULT 15
 LOCUS AV533632 61 bp mRNA linear EST 01-SEP-2000
 DEFINITION AV533632 Arabidopsis thaliana flower buds Columbia Arabidopsis
 thaliana cDNA clone FB065C05F 3', mRNA sequence.

ACCESSION AV533632
 VERSION AV533632
 KEYWORDS AV533632.1 GI:8693915
 EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 61)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)

JOURNAL MEDLINE
 PUBMED 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source Location/Qualifiers

1..61
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="FB065C05F"
 /tissue_type="flower buds"
 /clone_lib="Arabidopsis thaliana flower buds Columbia"
 /note="Vector: pBluescriptII SK-, Site_1: EcoRI, Site_2:
 XhoI"

BASE COUNT 21 a 4 c 12 g 24 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5.1e+03 Length: 61
 Score: 20.00 Matches: 4
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AV533632 (1-61)

OY 1 LysApGluLeu 4
 DB 9 AAAGATGATG 20

Search completed: August 26, 2003, 19:21:15
 Job time : 110.894 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:10:24 ; Search time 1733 Seconds
(without alignments) 8781.524 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372
Sequence: 1 aagcttaccatgcaggagta.....attccatcatgcctcgcag 372

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sbs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sbs: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rod: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_ey: *
39: em_higo_hum: *
40: em_higo_mus: *
41: em_higo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	372	100.0	372	6	AR121628
2	271.2	72.9	372	6	AR121631
3	264.8	71.2	372	6	AR121630
4	202	54.3	369	6	AR121627
5	194	52.2	369	6	AR121626
6	193.4	52.0	387	6	AR121623
7	192	51.6	387	6	AR121624
8	190.8	51.3	444	6	AR121632
9	177.6	47.7	357	6	AR121625
10	156.4	42.0	315	6	AR121629
11	142.6	38.3	2274	9	AB086984
12	142.6	38.3	2439	6	AX329932
13	142.6	38.3	2439	6	AX330698
14	142.6	38.3	2439	6	AX330917
15	142.6	38.3	2439	9	H0MCOMP
16	138	37.1	2710	9	BC033676
17	137.8	37.0	764	6	BD124595
18	137.8	37.0	764	6	BD126165
19	137.8	37.0	1779	9	BD127132
20	137.8	37.0	1779	9	AK074508
21	128.8	34.6	2302	4	AF325902
22	101.8	27.4	2438	10	AF033530
23	99.6	26.8	305	12	EVE291687
24	98.4	26.5	2421	10	RNCOMP
25	83.6	22.5	46275	9	AC003107
26	80	21.5	302	12	EVE291686
27	78.2	21.0	323	12	EVE291685
28	74	19.9	78	6	A92276
29	73.8	19.8	78	6	A92275
30	64.2	17.3	5719	10	MMCATS1
31	63.4	17.0	8923	10	AF318577
32	63.4	17.0	204302	2	AC068650
33	63.4	17.0	234787	2	AC073780
34	61	16.4	180286	10	AL645600
35	60.2	16.2	8524	10	AF257516
36	59.2	15.9	188880	10	AC093203
37	59.2	15.9	192939	2	AC112412
38	59.2	15.9	233181	2	AC111558
39	58.8	15.8	1440	14	BHV1UL7GN
40	58.6	15.8	225024	2	AC116415
41	58.4	15.7	161005	2	AC116595
42	58.4	15.7	172114	10	AC131903
43	58.2	15.6	261604	2	AC119819
44	58	15.6	3454	14	HAD293912
45	58	15.6	3454	14	HAD293913

ALIGNMENTS

RESULT 1
LOCUS AR121628
DEFINITION Sequence 24 from patent US 6160088.
ACCESSION AR121628
VERSION AR121628.1 GI:14105204
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 372)
AUTHORS Rochman,J.E., Mayhew,M. and Hoe,M.H.
TITLE KDEL receptor inhibitors
JOURNAL Patent: US 6160088-A 24 12-DEC-2000;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

```
source 1..372
/organism="unknown"
BASE COUNT 91 a 102 c 122 g 57 t
ORIGIN
Query Match 100.0%; Score 372; DB 6; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.3e-67;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTACCATGGGAAGATGATGATTTTAGGCTTGCCCTTGCGGAGCTGCGAC 60
DB 1 AAGCTTACCATGGGAAGATGATGATTTTAGGCTTGCCCTTGCGGAGCTGCGAC 60
QY 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTGTTCAAGACCTGGGCGCGAGATG 120
DB 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTGTTGAGGAGACTTTAAACGGGAGTTTC 120
QY 121 CTTGGGGAATCTGCAGAAACCAACGGCGCTGCGAGAGCTGCGGAACTGGCTGCGGAG 180
DB 121 CTTGGGGAATCTGCAGAAACCAACGGCGCTGCGAGAGCTGCGGAACTGGCTGCGGAG 180
QY 181 CAGGTCAAGGAGATCACTTCTGAAAAACAACGAGTGAAGTGAACGCGTGGGCGCG 240
DB 181 CAGGTCAAGGAGATCACTTCTGAAAAACAACGAGTGAAGTGAACGCGTGGGCGCG 240
QY 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGAAACCG 300
DB 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGAAACCG 300
QY 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTTAGCGCGCGGAGATTTCCAT 360
DB 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTTAGCGCGCGGAGATTTCCAT 360
QY 361 ATGCATCTCGAG 372
DB 361 ATGCATCTCGAG 372

RESULT 2
LOCUS AR121631 372 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 30 from patent US 6160088.
ACCESSION AR121631
VERSION AR121631.1 GI:14105207
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 372)
AUTHORS
Rothman,J.E., Mayhew,M. and Hoe,M.H.
TITLE
KDEL receptor inhibitors
JOURNAL
Patent: US 6160088-A 30 12-DEC-2000;
FEATURES
Location/Qualifiers
source 1..372
BASE COUNT 102 a 96 c 106 g 68 t
ORIGIN
Query Match 72.9%; Score 271.2; DB 6; Length 372;
Best Local Similarity 83.1%; Pred. No. 2.1e-46;
Matches 309; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```
QY 181 CAGGTCAAGGAGATCACTTCTGAAAAACAACGATGATGATGAGCGCGTGGCGCG 240
DB 181 CAGGTTAAGAAACATCATTTTGGGAAACCAATAGTGAATGCAAGCTTGGCGTCCG 240
QY 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGAAACCG 300
DB 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGAAACCG 300
QY 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTTAGCGCGCGGAGATTTCCAT 360
DB 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTTAGCGCGCGGAGATTTCCAT 360
QY 361 ATGCATCTCGAG 372
DB 361 ATGCATCTCGAG 372

RESULT 3
LOCUS AR121630 372 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 28 from patent US 6160088.
ACCESSION AR121630
VERSION AR121630.1 GI:14105206
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 372)
AUTHORS
Rothman,J.E., Mayhew,M. and Hoe,M.H.
TITLE
KDEL receptor inhibitors
JOURNAL
Patent: US 6160088-A 28 12-DEC-2000;
FEATURES
Location/Qualifiers
source 1..372
BASE COUNT 96 a 105 c 109 g 62 t
ORIGIN
Query Match 71.2%; Score 264.8; DB 6; Length 372;
Best Local Similarity 82.0%; Pred. No. 4.5e-45;
Matches 305; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
```

```
QY 1 AAGCTTACCATGGGAAGATGATGATTTTAGGCTTGCCCTTGCGGAGCTGCGAC 60
DB 1 AAGCTTACCATGGGAAGATGATGATTTTAGGCTTGCCCTTGCGGAGCTGCGAC 60
QY 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTGTTCAAGACCTGGGCGCGAGATG 120
DB 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTGTTGAGGAGACTTTAAACGGGAGTTTC 120
QY 121 CTTGGGGAATCTGCAGAAACCAACGGCGCTGCGAGAGCTGCGGAACTGGCTGCGGAG 180
DB 121 CTTGGGGAATCTGCAGAAACCAACGGCGCTGCGAGAGCTGCGGAACTGGCTGCGGAG 180
QY 181 CAGGTCAAGGAGATCACTTCTGAAAAACAACGAGTGAAGTGAACGCGTGGGCGCG 240
DB 181 CAGGTCAAGGAGATCACTTCTGAAAAACAACGAGTGAAGTGAACGCGTGGGCGCG 240
QY 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGAAACCG 300
DB 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGAAACCG 300
QY 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTTAGCGCGCGGAGATTTCCAT 360
DB 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTTAGCGCGCGGAGATTTCCAT 360
QY 361 ATGCATCTCGAG 372
DB 361 ATGCATCTCGAG 372

RESULT 4
LOCUS AR121627
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 372)
AUTHORS
Rothman,J.E., Mayhew,M. and Hoe,M.H.
TITLE
KDEL receptor inhibitors
JOURNAL
Patent: US 6160088-A 30 12-DEC-2000;
FEATURES
Location/Qualifiers
source 1..372
BASE COUNT
ORIGIN
```

LOCUS AR121627 369 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 22 from patent US 6160088.
 ACCESSION AR121627
 VERSION AR121627.1 GI:14105203
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 369)
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 22 12-DEC-2000;
 FEATURES
 source 1..369
 /organism="unknown"
 BASE COUNT 94 a 94 c 121 g 60 t
 ORIGIN

Query Match 54.3%; Score 202; DB 6; Length 369;
 Best Local Similarity 79.1%; Pred. No. 5.2e-32;
 Matches 239; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 71 AAGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 130
 DB 68 AGGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 127
 DB 131 TGCAGAAACCAACGCGCGCTGCGAGACGTCGGGAGACTGGGCGAGCGATGCGAGG 190
 DB 128 TAAACCAATGATGATGATGCTGGAGAGCTCCGAGATGTCATGAGACGCGGTGAAG 187
 QY 191 AGATCACGCTTCCTGAAAAACGCGTATGATGAGCGCTGGCGGCGCGAGCCGACG 250
 DB 188 AGACCATGTTCTTGAGAAACATTCAGATGCGAGCCCTGTGGCCGCGAGCCGACG 247
 QY 251 CGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 310
 DB 248 CGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 307
 QY 311 GTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATCTCG 370
 DB 308 GTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATCTCG 367
 QY 371 AG 372
 DB 368 AG 369

RESULT 5
 LOCUS AR121626 369 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 20 from patent US 6160088.
 ACCESSION AR121626
 VERSION AR121626.1 GI:14105202
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 369)
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 20 12-DEC-2000;
 FEATURES
 source 1..369
 /organism="unknown"
 BASE COUNT 87 a 103 c 120 g 59 t
 ORIGIN

Query Match 52.2%; Score 194; DB 6; Length 369;
 Best Local Similarity 77.5%; Pred. No. 2.4e-30;
 Matches 234; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 71 AAGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 130
 DB 68 AGGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 127
 DB 131 TGCAGAAACCAACGCGCGCTGCGAGACGTCGGGAGACTGGGCGAGCGATGCGAGG 190
 DB 128 TAAACCAATGATGATGATGCTGGAGAGCTCCGAGATGTCATGAGACGCGGTGAAG 187
 QY 191 AGATCACGCTTCCTGAAAAACGCGTATGATGAGCGCTGGCGGCGCGAGCCGACG 250
 DB 188 AGACCATGTTCTTGAGAAACATTCAGATGCGAGCCCTGTGGCCGCGAGCCGACG 247
 QY 251 CGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 310
 DB 248 CGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 307
 QY 311 GTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATCTCG 370
 DB 308 GTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATCTCG 367
 QY 371 AG 372
 DB 368 AG 369

DB 68 AGGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGCCAGC 127
 QY 131 TGCAGAAACCAACGCGCGCTGCGAGACGTCGGGAGACTGGGCGAGCGATGCGAGG 190
 DB 128 TCAACCTTTCAACCAAGATCTCTAGTGAGCTTGGGAGACATCCGAGACGAGTGAAG 187
 QY 191 AGATCACGCTTCCTGAAAAACGCGTATGATGAGCGCTGGCGGCGCGAGCCGACG 250
 DB 188 AATGTCACTATCCGGAACACATCATGATGATGATGATGATGATGATGATGATGAT 247
 QY 251 CGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 310
 DB 248 CGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 307
 QY 311 GTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATCTCG 370
 DB 308 GTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATCTCG 367
 QY 371 AG 372
 DB 368 AG 369

RESULT 6
 LOCUS AR121623 387 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 14 from patent US 6160088.
 ACCESSION AR121623
 VERSION AR121623.1 GI:14105199
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 387)
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
 TITLE KDEL receptor inhibitors
 JOURNAL Patent: US 6160088-A 14 12-DEC-2000;
 FEATURES
 source 1..387
 /organism="unknown"
 BASE COUNT 94 a 110 c 125 g 58 t
 ORIGIN

Query Match 52.0%; Score 193.4; DB 6; Length 387;
 Best Local Similarity 81.6%; Pred. No. 3.2e-30;
 Matches 249; Conservative 0; Mismatches 26; Indels 30; Gaps 1;

QY 98 GTTCAGACCTGGGCGCGAGATGCTTCGGAATGCGAGAAACCAACGCGCGCTGCAG 157
 DB 83 GTGAGACCTGAGCCCAACATGCTTGAGAACTCCAGAGACTAATGCGCGCTGCAG 142
 QY 158 ACGTGGGAGCTGGCTGCGGAGAGCTCAGGAGATCACTTCTGAAAAACGCGTGA 217
 DB 143 ACGTGGGAGAGCTGCTGCGGAGAGCTCAGGAGATCACTTCTGAAAAACGCGTGA 202
 QY 218 TGGAGTGTAGCGCGTGC-----GGCGGAGCGCG 247
 DB 203 TGGAGTGTAGCGCGTGC-----GGCGGAGCGCG 262
 QY 248 AGCGGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 307
 DB 263 AGCGGAAACCGCAGCGCGAGCGCGAGCGCGCGAAACCGCAGCGCGAAACCGGAACG 322
 QY 308 AAGTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATC 367
 DB 323 AAGTACCGGATCATCAGAAAAAGATGATGTAGCGGCGCGAGAAATTCATATGATC 382
 QY 368 TCGAG 372
 DB 383 TCGAG 387

RESULT 7

AR121624	AR121624	397 bp	DNA	linear	PAT 16-MAY-2001
LOCUS					
DEFINITION	Sequence 16 from patent US 6160088.				
ACCESSION	AR121624				
VERSION	AR121624.1	GI:14105200			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 387)				
TITLE	Rochman,J.E., Mayhew,M. and Hoe,M.H.				
JOURNAL	KDEL receptor inhibitors				
FEATURES	Patent: US 6160088-A 16-12-DEC-2000;				
	Location/Qualifiers				
Source	1..387				
	/organism="unknown"				
BASE COUNT	93 a 107 c 126 g 61 t				
ORIGIN					

Query Match	51.6%;	Score 192;	DB 6;	Length 387;
Best Local Similarity	76.4%;	Pred. No. 6.2e-30;		
Matches 307;	Conservative	0;	Mismatches 50;	Indels 45;
				Gaps 4;

QY 1 AAGCTTACCAATGGGAAAGGTATCATATTTTAAAGCTTGTGCGCTTGGGCGAGTCTGACG 60

Db 1 AAGCTTACCAATGGGAAATTCACTGTGT - GACGCGGCGCTTGGCTGTCTGGGCGCGCT 59

QY 61 GCTGCACAAAAAGATTCAGCGCTGGGTGGAGACTGTGTTGCAGACTTGGGCGCCAGATG 120

Db 60 GCGGGC - CGAGGGATTCAGCGCTGGGTGGAGACTGTGT ----- CACAGATG 105

QY 121 CTTGGGGAACTGCGAGAAACCAACGCGGCGCTGCAGACGTTGCGGGACTGCGTGGCAG 180

Db 106 CTTGGAGAACTTCGAGGAACTTAAATGCGGCGCTGCAGACGTAAGAGAGACTTTCGACAG 165

QY 181 CAGGTCAGGGAGATCACTGTTCTGTAACAAAACACGGGATGAGAGTGATGACGCTGGG ----- 235

Db 166 CAGGTCAGGGAGATCACTTCTGTAAGATACGGGATGAGAGTGATGAGCGCTTGGGAAATG 225

QY 236 ----- GGC CGCAGCGCGACCGGAAACCGCAGCCGACGCG 270

Db 226 CAGCCCGCAGCGACCCCGGTACTAAGTCCGACCGCAGCCGGAACCGCAGCCGACGCG 285

QY 271 CAGCCCGCAGCGGAAACCGCAGCCGGAACCGGAAACCGGAAAGGTACCGATCATCAGAAAA 330

Db 286 CAGCCCGCAGCGGAAACCGCAGCCGGAACCGGAAACCGGAAAGGTACCGGATCATCAGAAAA 345

QY 331 GATGAGTTGTATGGCGCGCGCGAGAAATTCATATGCATCTCGAG 372

Db 346 GATGAGTTGTATGGCGCGCGCGAGAAATTCATATGCATCTCGAG 387

RESULT B				
AR121632				
LOCUS	AR121632	444 bp	DNA	linear
DEFINITION	Sequence 35 from patent US 6160088.			PAT 16-MAY-2001
ACCESSION	AR121632			
VERSION	AR121632.1	GI:14105208		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified:			
AUTHORS	1 (bases 1 to 444)			
TITLE	Rochman,U.E., Mayhew,M. and Hoe,M.H.			
JOURNAL	KDEL receptor inhibitors			
FEATURES	Patent : US 6160088-A 35 12-DEC-2000;			
source	Location/Qualifiers			
	1..444			
	/organism="unknown"			
BASE COUNT	115 a 122 c 134 g 73 t			
ORIGIN				
Query Match	51.3%	Score 190.8;	DB 6;	Length 444;

	Best Local Similarity	80.6%;	Pred. No. 1,1e-29;	
	Matches 266;	Conservative 0;	Mismatches 22;	Indels 42; Gaps 2;
QY	73	GGATCCAGCCCTGGGTGAGACTGTTGTTTCAGACCTGGGCCCCGAGATGCTTCGGGAAC	132	
Db	127	GGATCCAGCCCTGGGTGAGACTGTTGTTTCAGACCTGGGCCCCGAGATGCTTCGGGAAC	174	
QY	133	CAGGAAACCAACGCGCGCTGCAGACAGTCGCGGACTGCTGCGGACAGCATCAGGAG	192	
Db	175	CAGGAGACTAATGGGGCGCTGCAGAGAGTAAAGAGCTCTTGGCAGACGACAGTCAAGAG	234	
QY	193	ATCAGCTTCTGAAAAACACGGTGATGAGTGTGACGCGTGGC-----	235	
Db	235	ATCAGCTTCTGAAAGATACGGTGATGAGATGTGAGCGCTTGGGAAATGACGCCGACGC	294	
QY	236	-----GGCGCGACGCGCGACGCCGAAACCGACGCCGACGCCGACGCCG	282	
Db	295	ACCCCGGACTAGTGTCCGACGCCGACGCCGAAACCGACGCCGACGCCGACGCCG	354	
QY	283	AAACCGGAGCCGAAACCGGAAACCGGAAGATACCGGATCATCAGAAAAAAGTGAATTG	342	
Db	355	AAACCGGAGCCGAAACCGGAAACCGGAAGATACCGGATCATCAGAAAAAAGTGAATTG	414	
QY	343	GCGGCGCAGAAATCCATATGACATTCGAG	372	
Db	415	GCGGCGCAGAAATCCATATGACATTCGAG	444	

[illegible]

	Query Match	74.7%;	Score 177.6;	DB 6;	Length 357;	
	Best Local Similarity	43.4%;	Pred. No. 6,2e-27;			
	Matches	273;	Conservative	0;	Mismatches	84; Indels 15; Gaps 3;
QY	1	AAGCTTACATGAGGAAGTACATGATTTTTAGCGTCCTGCCTCCGCACTGTSCAGC	60			
Db	1	AAGCTTACATGAGGAAGTACATGATGTGT-GGGCGGCGGCGTTGCTGTGTCGGCCGGT	59			
QY	61	GCTGCCAAAAAAGSATICAGCCTTGSGTGAGACTGTTGTTCAGAACCCTGGGCCCGCAGATG	120			
Db	60	GCGGGC--CGAGGGATTCAGCTCGGGGTGGAGACTGTTGTAAAGCATATGGTCAACCCAG---	114			
QY	121	CTTCGGSAATGCAAGGAACAACGGCGCGCTCACAGACGTGGGGGAACTGCTGCGCGAG	180			
Db	115	-----CTCACCTCTTTCAACAGATCTTAAGTGAAGCTTCGGGACGACATTCGAGAC	165			
QY	181	CAGGTCAGGAGATTCAGGTTCTCTGA AAAACACGSGTAGTGAGTGCAGCGCTGCGGGCCG	240			
Db	166	CAGGTGAAGAAATGTCATCTCATTCGGAMACCATATGAGTGTCAAGGTGTGCGGTCCG	225			
QY	241	CAGCCGACGCCGA AACCGCAGCCGACCGCAGCCGACCGCAAAACCGCAGCCGA AACC	300			
Db	226	CAGCCGACGCCGA AACCGCAGCCGACCGCAGCCGACCGCAAAACCGCAGCCGA AACC	285			
QY	301	GAAACCGGAAGTACCGGATCATCAGAAAAGATGATGTTGAGCGGCCGACGAAATTCAT	360			

```

Db      286 GAACCGGAAGGTACCGGATCATCAGAAAAAGTAGTGTAGGCGCGCGAGAAATTCAT 345
Oy      361 ATGCATCTCGAG 372
Db      346 ATGCATCTCGAG 357

RESULT 10
LOCUS   AR121629 315 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 26 from patent US 6160088.
ACCESSION AR121629
VERSION   AR121629.1 GI:14105205
KEYWORDS
SOURCE   Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 315)
AUTHORS  Rothman,J.E., Mayhew,M. and Hoe,M.H.
TITLE     KDEL receptor inhibitors
JOURNAL   Patent: US 6160088-A 26 12-DEC-2000;
FEATURES  Location/Qualifiers
           source          1..315
                        /organism="unknown"
BASE COUNT  81 a 85 c 79 g 70 t
ORIGIN
Query Match 42.0%; Score 156.4, DB 6; Length 315;
Beat Local Similarity 71.0%; Pred. No. 1.6e-22;
Matches 264; Conservative 0; Mismatches 51; Indels 57; Gaps 2;

Oy      1 AAGCTTACATGGGAAGGATGATGATTTAGGCTTCTCGCCCTTGGCGGACATCTGCAGC 60
Db      1 AAGCTTACATGGGAAGGATGATGATTTAGGCTTCTCGCCCTTGGCGGACATCTGCAGC 60
Oy      61 GCTGCCAAAAAGATGATCAGCTGGGTGGAGACTGTGTTTCAGACTGGGCGCGGAGATG 120
Db      61 GCTGCCAAAAAGATGATCAGCTGGGTGGAGACTGTGTTTCAGAAAA----- 104
Oy      121 CTTCGGGAATGCGAGAAACCAACGCGGCTGCGAGACGTGCGGACTGCTGCGGCGAG 180
Db      105 -----GCTACGAATCTATTATCAATTTCTGTCTC 135
Oy      181 CAGGTCAAGGAGATGACGTTCTCTGAAAAACACGGTGTGATGATGACGCGCGGCGCG 240
Db      136 ATCTTAATATGCTCTTCTGATCTGTATCATCGTAATGCTTCTC-----CCG 183
Oy      241 CAGCGCAGCCGAAACCGCAGCGCGCAGCGCAGCGGAAACCGCAGCGGAAACCG 300
Db      184 CAGCGCAGCCGAAACCGCAGCGCGCAGCGCGCAGCGGAAACCGCAGCGGAAACCG 243
Oy      301 GAACCGGAAGGTACCGGATCATCAGAAAAAGTAGTGTAGGCGCGCGAGAAATTCAT 360
Db      244 GAACCGGAAGGTACCGGATCATCAGAAAAAGTAGTGTAGGCGCGCGAGAAATTCAT 303
Oy      361 ATGCATCTCGAG 372
Db      304 ATGCATCTCGAG 315

RESULT 11
LOCUS   AB086984 2274 bp mRNA linear PRI 07-JAN-2003
DEFINITION Homo sapiens comp mRNA for cartilage oligomeric matrix protein,
complete cds.
ACCESSION AB086984
VERSION   AB086984.1 GI:27530065
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.

```

```

REFERENCE 1
AUTHORS  Hashimoto,Y. and Mori,H.
TITLE     Human comp cDNA with 5 SNIPS
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 2274)
AUTHORS  Hashimoto,Y. and Mori,H.
TITLE     Direct Submission
JOURNAL   Submitted (27-JUN-2002) Hiroshi Mori, Osaka City University Medical
School, Neuroscience; 1-4-3 Asahimachi, Abeno-ku, Osaka 545-8585,
Japan (E-mail: moriemed.osaka-cu.ac.jp, Tel: 81-6-6645-3920,
Fax: 81-6-6645-3922)
FEATURES  Location/Qualifiers
           source          1..2274
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /cell_type="fibroblast"
                        1..2274
                        /gene="comp"
                        1..2274
                        /gene="comp"
                        /codon_start=1
                        /product="cartilage oligomeric matrix protein"
                        /protein_id="BAC53888.1"
                        /db_xref="GI:27530066"
                        /translation="MVPDTACVLLTLTAAAGASGGGSPGSDIGPQMLREJORTNAA
                        LODVDMRLROOVREITFLKNTVMEDACAGQOQSVRTGLPSVRPLLHGAPGCFPGVAC
                        IQTSGGRCGCPNAGFTNGSHCTDVANECNAHPCFPYRCLNTSPGRCEACPPGYSYG
                        PTHGGVGLAEFKAKQVCTDINBCTGQHNKCVNSVCINRGSPQCGPCQGFVGDQA
                        SGCGRGARPCPCPSBCEHADVLERDGRSCVCRVAGAGNGLICGRPTDIDGFP
                        DEKRCPEPCQKNDKCVTPNSGQEDVDRIIDGACPDADGDEVPKNDKCPVLRNP
                        DQRTDDBDKMGDADCNCRSOKNDQKDTDODGRGDACDDIDGDIRINQADNCPVRNP
                        SPQSDSDSDGIGDAGCNCPKSNPDQADVDHDFYGDACDSDPDODGPGHDSRNDCTP
                        VPSAQSDBDGGDADDDDDNDGCPDSDNKRLLVPNPGQENADBDGVDYQDDP
                        DADRVVKKIDVCPENAEVTLTDFRAFTVVLDEPGDAQIDENWVLLNQREIVTQWS
                        DPGLAVGYTLANGVDPEGTFRHNTVTDVDAAGFLFGYQDSSFYVMMKQMEQYQWA
                        NPFAVAPEGIQLAAVKSSTGPEQLNMLAHTGDTESQVRLWKDKDPRNVGMKDKSY
                        RMFLQHRPQVGYIRVFRYEGPELVADSNVVLDTTRMGRLGVCFCSQENIIMANLRYR
                        CNDRIIPDYERHQLRQA"
                        150
                        /gene="comp"
                        /note="SNIPS of # 1, 2, 3 are localized at the pentamer
                        formation domain"
                        /replace="g"
                        151
                        /gene="comp"
                        /note="SNIPS of # 1, 2, 3 are localized at the pentamer
                        formation domain"
                        /replace="c"
                        152
                        /gene="comp"
                        /note="SNIPS of # 1, 2, 3 are localized at the pentamer
                        formation domain"
                        /replace="t"
                        326
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"
                        /replace="c"
                        327
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"
                        /replace="g"
                        669
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"
                        /replace="g"
                        670
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"

```



```

DEPLAAVGYTA.FNCLVDPEGTGTHVNTVDDVDYAGFLFGYQDSS.FVYVMMKKMEORTYVWQ
NPPRAVAABEIGTQAKAYKSS.TGTPBQRLNALMHTGDETSQYRILMKDRANCKRKSSY
RWFLOHRPOVGYI.RVRFYBESPELVADSNVLDITTMGRGLGVCFSESENTIIMANLRFS
CNDIPEDYETHQLRQA"
26. .85
/gene="COMP"
290. .892
/ncore="putative"
/rpt_family="thrombospondin type 2"
/rpt_type=tandem
893. .1577
/ncore="putative"
/rpt_family="thrombospondin type 3"
/rpt_type=tandem
2420. .2425
/gene="COMP"
2439
/gene="COMP"

polyA_site
BASE COUNT      503 a      758 c      809 g      369 t
ORIGIN
Query Match      38.3%; Score 142.6; DB 9; Length 2439;
Best Local Similarity 97.3%; Pred. No. 9.9e-20;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 98 GTTCAGACCTGGGACCCGACATGCTTGGGGAATGACAGAAACCAACGCGCGCTGCAGG 157
Db 105 GCTCAGACCTGGGACCCGCAATGCTTGGGGAATGACAGAAACCAACGCGCGCTGCAGG 164
QY 158 ACGTGGCGGACTGGAGTCGCGCAGCAGAGGTGACGGAGATCACGTTCTGAAAAAACAAGGTGA 217
Db 165 ACGTGGCGGACTGGAGTCGCGCAGCAGAGGTGACGGAGATCACGTTCTGAAAAAACAAGGTGA 224
QY 218 TGGAGTGTGACGCGTGGCGGCGCGGACGCG 246
Db 225 TGGAGTGTGACGCGTGGCGGCGGATGACAGCAG 253

```

11 slide Page Blank (usps)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:10:25 ; Search time 207 Seconds

(without alignments)
4851.164 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372
Sequence: 1 aagcttaccatgggaagta.....attccatcgcctcgag 372

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.GeneSeq_19Jun03.*

1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	357	96.0	369	21	AAZ50497	KOEL receptor inh1
2	256.2	68.9	21	AAZ50500		KOEL receptor inh1
3	249.8	67.2	369	21	AAZ50499	KOEL receptor inh1
4	202	54.3	369	21	AAZ50496	KOEL receptor inh1
5	194	52.2	369	21	AAZ50495	KOEL receptor inh1
6	193.4	52.0	387	21	AAZ50492	KOEL receptor inh1
7	192	51.6	387	21	AAZ50493	KOEL receptor inh1
8	190.8	51.3	444	21	AAZ50501	KOEL receptor inh1

9	177.6	47.7	357	21	AAZ50494
10	142.6	38.3	755	21	AAA47734
11	142.6	38.3	925	21	AAA47735
12	142.6	38.3	2439	24	ABT10887
13	142.6	38.3	2439	24	ABT07751
14	142.6	38.3	2439	24	ABL62104
15	142.6	38.3	2439	24	ABL62870
16	142.6	38.3	2439	24	ABL63089
17	142.6	38.3	2439	25	ACC50112
18	142.6	38.3	2439	25	ABX76334
19	141.4	38.0	312	21	AAZ50498
20	137.8	37.0	764	22	AAK91566
21	137.8	37.0	764	22	AAK91336
22	137.8	37.0	1779	22	AAK94103
23	83.6	22.5	8532	22	ABA15407
24	83.6	22.5	8532	22	ABA15585
25	83.6	22.5	46275	24	ABT10145
26	73.8	19.8	78	19	AAV27572
27	58	15.6	34427	20	AAK07371
28	58	15.6	34616	24	ABK49009
29	58	15.6	35937	24	ABK69881
30	58	15.6	36335	15	AAK68003
31	56.4	15.2	480	17	AAK41508
32	55	14.8	1037	21	AAA59242
33	55	14.8	1159	21	AAA59240
34	55	14.8	1472	21	AAA59241
35	53.8	14.5	397	20	ABK69891
36	53.6	14.4	403	21	ABK10240
37	53.6	14.4	5120	22	AAK64677
38	53.4	14.4	1865	23	ABL03153
39	53.4	14.4	5178	23	ABL03152
40	53	14.2	521	21	AAK66020
41	53	14.2	521	24	ABO92425
42	53	14.2	521	24	ABL49235
43	53	14.2	1082	21	AACT8047
44	53	14.2	1412	24	ABK64787
45	53	14.2	3747	25	ACA03912

ALIGNMENTS

RESULT 1
AAZ50497
ID AAZ50497 standard; DNA; 369 BP.

AAZ50497;

23-MAY-2000 (first entry)

KDEL receptor inhibitor-6 DNA.

KDEL receptor inhibitor; heat shock protein; immune response;

oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;

melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

infectious disease; allergy; autoimmune disease; ss.

Chimeric - Adenovirus E3.

Chimeric - Homo sapiens.

Chimeric - Camelus sp.

Key CDS

Location/Qualifiers

10..339

/*tag= a

/product= "KDEL receptor inhibitor protein"

/*tag= b

/note= "Derived from adenovirus E3"

/*tag= c

/product= "Mature KDEL receptor inhibitor protein"

97..234

/*tag= d

KDEL receptor inh
Human COMP/TSP-1 c
Human COMP/TSP-2 c
Human breast cance
Breast cancer-asso
Colon adenocarcino
Breast cancer rela
Breast cancer rela
Breast cancer asso
Lung cancer-associ
KDEL receptor inh
Human cDNA 5'-end
Human cDNA clone x
Human full-length
Human nervous syst
Human nervous syst
Human breast cance
Nucleotide sequenc
Adenovirus vector
Adenoviral vector
Human adeno-associ
Ad2/-ORF6/PGK-CFTR
Human adenovirus t
Exons E, C and A o
An EcoRI fragment
Exons D, C, B and
Spincerebellar at
Trinucleotide repe
DNA sequence of hu
Drosophila melanog
Drosophila melanog
Human lung cancer-
Human lung cancer
Human lung tumor
Human cancer assoc
Human cDNA differe
cDNA downregulated

```

FT      /note= "Human COMP pentamerisation domain"
FT      misc_feature      235..306
FT      /*tag= e
FT      /note= "Camel IgG linker domain"
XX
XX      WO200006729-A1.
XX
XX      10-FEB-2000.
XX
XX      28-JUL-1999; 99WO-US17147.
XX
XX      29-JUL-1998; 98US-0124671.
XX
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX      Rothman JE, Mayhew M, Hoe MH;
XX      WPI: 2000-195296/17.
XX      P-PSDB; AMY44963.
XX
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
XX      PT domain useful for promoting secretion of proteins which are normally
XX      PT retained within the cell -
XX
XX      Disclosure; Fig 6; 87pp; English.
XX
XX      The patent discloses the use of KDEL receptor inhibitor to promote
XX      CC secretion of proteins that are normally retained within the cell such as
XX      CC heat shock proteins by inhibiting KDEL receptor-mediated return of
XX      CC protein complexes to endoplasmic reticulum. This makes the secreted heat
XX      CC shock proteins more accessible to the immune system and improves immune
XX      CC response to a target antigen. The inhibitor protein comprises several
XX      CC subunits where each subunit comprises an oligomerisation domain and has
XX      CC at its carboxy terminus a region which binds to a KDEL receptor. The
XX      CC target antigen may be associated with diseases including neoplasia such
XX      CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX      CC astrocytoma, with defective tumour suppressor genes, oncogenes,
XX      CC infectious diseases, allergy or autoimmune diseases. The present
XX      CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
XX      CC cleavable signal peptide; the oligomerisation domain of human cartilage
XX      CC oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG
XX      CC linker domain and the carboxy-terminal sequence KDEL. The subsequence
XX      CC GDCC is an alteration of rat COMP which provides increased stability via
XX      CC disulphide bonds.
XX
XX      Sequence 369 BP; 90 A; 102 C; 120 G; 57 T; 0 other;
XX
XX      Query Match      96.0%; Score 357; DB 21; Length 369;
XX      Best Local Similarity 100.0%; Pred. No. 6.2e-76;
XX      Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      16 AGGTACATGATTTTGGCTTGCCTGCGCGGAGTGTGAGCGGCTGCCAAAAAAGGA 75
XX      |||||||
XX      Db 13 AGGTACATGATTTTGGCTTGCCTGCGCGGAGTGTGAGCGGCTGCCAAAAAAGGA 72
XX
XX      76 TCCAGCTGGGTGAGACTGTTGTTGACACTGCGGCCCGCAGATGCTTGGGAACTGCAG 135
XX      |||||||
XX      Db 73 TCCAGCTGGGTGAGACTGTTGTTGACACTGCGGCCCGCAGATGCTTGGGAACTGCAG 132
XX
XX      136 GAAACCAAGCGCGCTGTGAGAGCGTGGCGGACTGGCTCGGCGAGAGTCAAGGAGATC 195
XX      |||||||
XX      Db 133 GAAACCAAGCGCGCTGTGAGAGCGTGGCGGACTGGCTCGGCGAGAGTCAAGGAGATC 192
XX
XX      196 AGCTTCTGTAACCAACGCGTGTGAGTGTGACGCGTGGCGGCGGAGCGGAGCCGANA 255
XX      |||||||
XX      Db 193 AGCTTCTGTAACCAACGCGTGTGAGTGTGACGCGTGGCGGCGGAGCGGAGCCGANA 252
XX
XX      256 CCGCAGCCGCGAGCCGCGAGCCGAGCCGAAACCGCAGCCGAAACCGGAAACCGAAGGTACC 315
XX      |||||||
XX      Db 253 CCGCAGCCGCGAGCCGCGAGCCGAGCCGAAACCGCAGCCGAAACCGGAAACCGAAGGTACC 312
XX
XX      316 GGATCATCGAAAAAGATGATGTTGTTAGGGCGCGCGAATTCATATGTCATCTCCAG 372
XX      |||||||

```

```

Db      313 GGATCATCGAAAAAGATGATGTTGAGCGCGCGCAGAAATTCATATGTCATCTCCAG 369
XX
XX      RESULT 2
XX      ID      AA250500 standard; DNA; 369 BP.
XX
XX      AA250500;
XX
XX      23-MAY-2000 (first entry)
XX
XX      KDEL receptor inhibitor-9 DNA.
XX
XX      KDEL receptor inhibitor; heat shock protein; immune response;
XX      KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX      KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX      KM infectious disease; allergy; autoimmune disease; ss.
XX
XX      Chimeric - Adenovirus E3.
XX      OS Chimeric - Homo sapiens.
XX      OS Chimeric - Camelus sp.
XX
XX      Key      Location/Qualifiers
XX      CDS      10..339
XX      FT      /*tag= a
XX      FT      /product= "KDEL receptor inhibitor protein"
XX      FT      10..69
XX      FT      /*tag= b
XX      FT      /note= "Derived from adenovirus E3"
XX      FT      70..336
XX      FT      /*tag= c
XX      FT      /note= "Mature KDEL receptor inhibitor protein"
XX      FT      97..234
XX      FT      /*tag= d
XX      FT      /note= "Human TSP4 trimerisation domain"
XX      FT      235..306
XX      FT      /*tag= e
XX      FT      /note= "Camel IgG linker domain"
XX
XX      WO200006729-A1.
XX
XX      10-FEB-2000.
XX
XX      28-JUL-1999; 99WO-US17147.
XX
XX      29-JUL-1998; 98US-0124671.
XX
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX      Rothman JE, Mayhew M, Hoe MH;
XX      WPI: 2000-195296/17.
XX      P-PSDB; AMY44966.
XX
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
XX      PT domain useful for promoting secretion of proteins which are normally
XX      PT retained within the cell -
XX
XX      Disclosure; Fig 9; 87pp; English.
XX
XX      The patent discloses the use of KDEL receptor inhibitor to promote
XX      CC secretion of proteins that are normally retained within the cell such as
XX      CC heat shock proteins by inhibiting KDEL receptor-mediated return of
XX      CC protein complexes to endoplasmic reticulum. This makes the secreted heat
XX      CC shock proteins more accessible to the immune system and improves immune
XX      CC response to a target antigen. The inhibitor protein comprises several
XX      CC subunits where each subunit comprises an oligomerisation domain and has
XX      CC at its carboxy terminus a region which binds to a KDEL receptor. The
XX      CC target antigen may be associated with diseases including neoplasia such
XX      CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX      CC astrocytoma, with defective tumour suppressor genes, oncogenes,
XX      CC infectious diseases, allergy or autoimmune diseases. The present
XX      CC sequence encodes KDEL receptor inhibitor comprising regions encoding a

```

CC cleavable signal peptide; the oligomerisation domain of human
 CC thrombospondin 4 (TSP4) trimerisation domain; a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an
 CC alteration of rat cartilage oligomeric matrix protein which provides
 CC increased stability via disulphide bonds.

XX Sequence 369 BP; 101 A; 96 C; 104 G; 68 T; 0 other;

Query Match 68.9%; Score 256.2; DB 21; Length 369;

Best Local Similarity 82.4%; Pred. No. 6,8e-52; Mismatches 63; Indels 0; Gaps 0;

Matches 294; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 16 AGGTACATGATTTTGGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGGA 75
 DB 13 AGGTACATGATTTTGGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGGA 72
 QY 76 TCCAGCTGGGTGAGACTGTTGTTCAAGCTGGGCCCGCAGATCTTCGGAACTGCAG 135
 DB 73 TCCAGCTGGGTGAGACTGTTGTTCAAGCTGGGCCCGCAGATCTTCGGAACTGCAG 132
 QY 136 GAAACCAAGCGCGCGCTGACAGCTGGGGGAGCTGGCGAGCGTCCAGGAGATC 195
 DB 133 CAATTAAACCACTCTTGAGAGAGGTGAAGACCTTCTGAGACACAGGTTAAGAAACA 192
 QY 196 ACGTTCCTGAAAAACACCGGTGATGAGTGTGACCGGTGCGGGCCGACCGCAGCCGAAA 255
 DB 193 TCATTTTTGGCGAAACACCATAGCTGAATGCCAGCTTGCGGTCCCGACCGCAGCCGAAA 252
 QY 256 CCGCAGCCGCGACCGCGACCGCGAAGCCGAAACCGGAAACCGGAAAGTACC 315
 DB 253 CCGCAGCCGCGACCGCGACCGCGAAGCCGAAACCGGAAACCGGAAAGTACC 312
 QY 316 GGATCATCAGAAAAAGATGATGATGAGCGCGCGCGAGAAATTCATATGATCTCGAG 372
 DB 313 GGATCATCAGAAAAAGATGATGATGAGCGCGCGAGAAATTCATATGATCTCGAG 369

RESULT 3

AAZ50499 standard; DNA; 369 BP.

XX AAZ50499;

XX 23-MAY-2000 (first entry)

XX KDEL receptor inhibitor-8 DNA.

XX KDEL receptor inhibitor; heat shock protein; immune response;

XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;

XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

XX infectious disease; allergy; autoimmune disease; ss.

XX Chimeric - Adenovirus E3.

XX Chimeric - Homo sapiens.

XX Chimeric - Camelus sp.

XX Key Location/Qualifiers

FT CDS 10..339

FT sig_peptide 10..69

FT mat_peptide 70..336

FT misc_feature 97..234

FT misc_feature 235..306

XX /note= "Camel IgG linker domain"

PN WO200006729-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17147.

XX 29-JUL-1998; 98US-0124671.

XX (SLOC) SLOAN KETTERING INST CANCER RES.

XX Rothman JE, Mayhew M, Hoe MH;

XX WPI; 2000-195296/17.

XX P-PSDB; AAY44965.

PT Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell

PS Disclosure; Fig 8; 87pp; English.

CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of human
 CC thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an
 CC alteration of rat cartilage oligomeric matrix protein which provides
 CC increased stability via disulphide bonds.

XX Sequence 369 BP; 95 A; 105 C; 107 G; 62 T; 0 other;

XX Query Match 67.2%; Score 249.8; DB 21; Length 369;

XX Best Local Similarity 81.2%; Pred. No. 2.3e-50; Mismatches 67; Indels 0; Gaps 0;

XX Matches 290; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 16 AGGTACATGATTTTGGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGGA 75
 DB 13 AGGTACATGATTTTGGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGGA 72
 QY 76 TCCAGCTGGGTGAGACTGTTGTTCAAGCTGGGCCCGCAGATCTTCGGAACTGCAG 135
 DB 73 TCCAGCTGGGTGAGACTGTTGTTCAAGCTGGGCCCGCAGATCTTCGGAACTGCAG 132
 QY 136 GAAACCAAGCGCGCGCTGACAGCTGGGGGAGCTGGCGAGCGTCCAGGAGATC 195
 DB 133 CAATTAAACCACTCTTGAGAGAGGTGAAGACCTTCTGAGACACAGGTTAAGAAATG 192
 QY 196 ACGTTCCTGAAAAACACCGGTGATGAGTGTGACCGGTGCGGGCCGACCGCAGCCGAAA 255
 DB 193 TCATTTCATCCGAAACACCATAGCTGAATGCCAGCTTGCGGTCCCGACCGCAGCCGAAA 252
 QY 256 CCGCAGCCGCGACCGCGACCGCGAAGCCGAAACCGGAAACCGGAAAGTACC 315
 DB 253 CCGCAGCCGCGACCGCGACCGCGAAGCCGAAACCGGAAACCGGAAAGTACC 312
 QY 316 GGATCATCAGAAAAAGATGATGATGAGCGCGCGCGAGAAATTCATATGATCTCGAG 372
 DB 313 GGATCATCAGAAAAAGATGATGATGAGCGCGCGAGAAATTCATATGATCTCGAG 369

RESULT 4

AAZ50496

ID AA250496 standard; DNA; 369 BP.
 XX
 AC AA250496;
 DT 23-MAY-2000 (first entry)
 XX
 DE KDEL receptor inhibitor-5 DNA.
 XX
 KM KDEL receptor inhibitor; heat shock protein; immune response;
 KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KM infectious disease; allergy; autoimmune disease; ss.
 XX
 OS Chimeric - Mus sp.
 OS Chimeric - Xenopus sp.
 OS Chimeric - Camelus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 10..339
 FT /tag= a
 FT /product= "KDEL receptor inhibitor protein"
 FT 10..68
 FT /tag= b
 FT /note= "Derived from mouse Bip"
 FT 69..336
 FT /tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"
 FT 97..234
 FT /tag= d
 FT /note= "Xenopus TSP4 trimerisation domain"
 FT 235..306
 FT /tag= e
 FT /note= "Camel IgG linker domain"
 XX
 FT WO200006729-A1.
 XX
 PD 10-FEB-2000.
 XX
 PF 28-JUL-1999; 99WO-US17147.
 XX
 PR 29-JUL-1998; 98US-0124671.
 XX
 PA (SLOK) SLOAN KETERING INST CANCER RES.
 XX
 PI Rothman JE, Mayhew M, Hoe MH;
 XX
 DR WPI: 2000-195296/17.
 DR P-PSDB: AAY44962.
 XX
 FT Inhibitors of the KDEL receptor which comprises an oligomerisation
 FT domain useful for promoting secretion of proteins which are normally
 FT retained within the cell -
 XX
 PS Disclosure: Fig 5; 87pp; English.
 XX
 CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of Xenopus
 CC thrombospondin 4 (TSP4) trimerisation domain including an additional
 CC sub-sequence; a camel IgG linker domain and the carboxy-terminal sequence
 CC KDEL. The subsequence GDCG is an alteration of rat cartilage oligomeric
 CC matrix protein which provides increased stability via disulphide bonds.
 CC

XX Sequence 369 BP; 94 A; 94 C; 121 G; 60 T; 0 other;
 SQ
 Query Match 54.3%; Score 202; DB 21; Length 369;
 Best Local Similarity 79.1%; Pred. No. 5.8e-39;
 Matches 239; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
 OY 71 AAGATCCAGCCTGGGTGAGACTGTGTTGACAGCTTGCGCCGCGAGTGTTCGGGAAC 130
 DB 68 AGGATCCAGCCTGGGTGAGACTGTGTTGAGCTGACGACGACAGTGTGATGGCCAGA 127
 OY 131 TGCAGAAACCAACCGCGCGCTGCAGACGTCGCGGACTGTGCTGGCGACAGCTCAAGG 190
 DB 128 TAACCCAAATGATCAGATGCTGGGAGACTCCGAGATGCATGAGACGAGGTGAAG 187
 OY 191 AGATCAGCTTCCTGAAACACGAGTGAATGAGTGAACGCTGCGCGCGCGACCGCAGC 250
 DB 188 AGACATGTTCTTGAAACACATTGACGATGCGACGCTGTGCGCCGACCGCGACG 247
 OY 251 CGAAACCGCAGCGCGACCGCGACCGCGACCGGAAACCGCAGCCGAAACCGGAAG 310
 DB 248 CGAAACCGCAGCGCGACCGCGACCGCGACCGGAAACCGGAAACCGGAAG 307
 OY 311 GTACCGGATCATCAAAAAGATGATGTTGAGCGCGCGCGAGAAATTCATATGATCTCG 370
 DB 308 GTACCGGATCATCAAAAAGATGATGTTGAGCGCGCGCGAGAAATTCATATGATCTCG 367
 OY 371 AG 372
 DB 368 AG 369
 RESULT 5
 AA250495
 ID AA250495 standard; DNA; 369 BP.
 XX
 AC AA250495;
 DT 23-MAY-2000 (first entry)
 XX
 DE KDEL receptor inhibitor-4 DNA.
 XX
 KM KDEL receptor inhibitor; heat shock protein; immune response;
 KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KM infectious disease; allergy; autoimmune disease; ss.
 XX
 OS Chimeric - Mus sp.
 OS Chimeric - Camelus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 10..339
 FT /tag= a
 FT /product= "KDEL receptor inhibitor protein"
 FT 10..68
 FT /tag= b
 FT /note= "Derived from mouse Bip"
 FT 69..336
 FT /tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"
 FT 97..234
 FT /tag= d
 FT /note= "Mouse TSP3 domain"
 FT 235..306
 FT /tag= e
 FT /note= "Camel IgG linker domain"
 XX
 PN WO200006729-A1.
 XX
 PD 10-FEB-2000.
 XX
 PF 28-JUL-1999; 99WO-US17147.
 XX

XX	29-JUL-1998;	98US-0124671.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.	
PI	Rochman JE, Mayhew M, Hoe MH;	
XX	WPI; 2000-199296/17.	
DR	P-PSDB; AAY44961.	
XX	Inhibitors of the KDEL receptor which comprises an oligomerization	
PT	domain useful for promoting secretion of proteins which are normally	
PT	retained within the cell -	
XX		
PS	Disclosure; Fig 4; 87pp; English.	
XX		
CC	The patent discloses the use of KDEL receptor inhibitor to promote	
CC	secretion of proteins that are normally retained within the cell such as	
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of	
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat	
CC	shock proteins more accessible to the immune system and improves immune	
CC	response to a target antigen. The inhibitor protein comprises several	
CC	subunits where each subunit comprises an oligomerization domain and has	
CC	at its carboxy terminus a region which binds to a KDEL receptor. The	
CC	target antigen may be associated with diseases including neoplasia such	
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and	
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,	
CC	infectious diseases, allergy or autoimmune diseases. The present	
CC	sequence encodes KDEL receptor inhibitor comprising regions encoding a	
CC	cleavable signal peptide; the oligomerization domain of mouse	
CC	thrombospondin 3 (TSP3) trimerisation domain including an additional	
CC	sub-sequence; a camel IgG linker domain and the carboxy-terminal sequence	
CC	KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric	
CC	matrix protein which provides increased stability via disulphide bonds.	
XX		
SO	Sequence 369 BP; 87 A; 103 C; 120 G; 59 T; 0 other;	
	Query Match	52.2%; Score 194; DB 21; Length 369;
	Best Local Similarity	77.5%; Pred. No. 4.7e-37;
	Matches 234; Conservative	0; Mismatches 68; Indels 0; Gaps 0;
QY	71 AAGATTCAGCGCTGGTGGAGACTGTGTCAGACTGGCCCGGAGATGCTTCGGGAAC	130
DB	68 AGGGATTCAGCGCTGGTGGAGACTGTGTTGGGGAGACCAAGGCATTGGTCACCCAGC	127
QY	131 TGCAGGAACCAACGCGCGCTGCAGAGACGTGCGGGAAGTGGTGGCGAGAGTCAAGG	190
DB	128 TCACCCCTTTCAACCAAGATCCTAGTGGAGCTTCGGAGCAACATCCGAGCCAGGTGAAG	187
QY	191 AGATCACGTTCTGTAAAAACAAGGTAGTAGTGTGACGCGTGCAGCGCGCAGCCGAC	250
DB	188 AAATGTACTACTATCCGGAACACATCATGGAATGTGAGTGTGGGTCCGAGCCGCGAC	247
QY	251 CGAAACCGCAGCCGACGCGCAGCCGACGCGGAAACCGAGCCGAAACCGGAACG	310
DB	248 CGAAACCGCAGCCGACGCGCAGCCGACGCGGAAACCGAGCCGAAACCGGAACG	307
QY	311 GTACCGGATCATCAAAAAAGTAGTGTAGCGCGCGGCAAAATTCATATGATCATTCG	370
DB	308 GTACCGGATCATCAAAAAAGTAGTGTAGCGCGCGGCAAAATTCATATGATCATTCG	367
QY	371 AG 372	
DB	368 AG 369	
RESULT 6		
AAZ50492		
XX	AAZ50492 standard; DNA; 387 BP.	
XX	AAZ50492;	
XX	23-MAY-2000 (first entry)	

DE	KDEL receptor inhibitor-1 DNA.
XX	
KW	KDEL receptor inhibitor; heat shock protein; immune response;
RW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KX	infectious disease; allergy; autoimmune disease; ss.
XX	
OS	Chimeric - Mus sp.
OS	Chimeric - Camelus sp.
OS	Chimeric - Rattus sp.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	/tag= a
FT	/product= "KDEL receptor inhibitor protein"
FT	sig_peptide
FT	/tag= b
FT	/note= "Derived from mouse Bip"
FT	mat_peptide
FT	/tag= c
FT	/product= "Mature KDEL receptor inhibitor protein"
FT	misc_feature
FT	/tag= d
FT	/note= "Rat COMP domain"
FT	misc_feature
FT	/tag= e
FT	/note= "Camel IgG linker domain"
XX	
PN	WO20006729-A1.
XX	
PD	10-FEB-2000.
XX	
PF	28-JUL-1999; 99WC-US17147.
PR	29-JUL-1998; 98US-0124671.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.
XX	
PI	Rothman JF, Mayhew M, Hoe MH;
DR	WPI; 2000-195296/17.
XX	P-PSDB; AAY44958.
PT	Inhibitors of the KDEL receptor which comprises an oligomerization
PT	domain useful for promoting secretion of proteins which are normally
XX	retained within the cell
XX	-
PS	Disclosure; Fig 1; 87pb; English.
XX	
CC	The patent discloses the use of KDEL receptor inhibitor to promote
CC	secretion of proteins that are normally retained within the cell such as
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat
CC	shock proteins more accessible to the immune system and improves immune
CC	response to a target antigen. The inhibitor protein comprises several
CC	subunits where each subunit comprises an oligomerisation domain and has
CC	at its carboxy terminus a region which binds to a KDEL receptor. The
CC	target antigen may be associated with diseases including neoplasia such
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,
CC	infectious diseases, allergy or autoimmune diseases. The present
CC	sequence encodes KDEL receptor inhibitor comprising regions encoding a
CC	cleavable signal peptide; the oligomerisation domain from rat cartilage
CC	oligomeric matrix protein; a camel IgG linker domain and the carboxy
CC	-terminal sequence KDEL. This is introduced into host cells by suitable
CC	vectors.
SQ	Sequence 387 BP; 94 A; 110 C; 125 G; 58 T; 0 other;
Query Match	52.0%; Score 193.4; DB 21; Length 387;
Best Local Similarity	81.6%; Pred. No. 6.6e-37;
Matches 249; Conservative	0; Mismatches 26; Indels 30; Gaps 1;

QY 98 GTTCAGACCTGGGCCCCGACGATGCTTGGGAACTGCAGAAACCGCGCGCTGCAG 157
DB 83 GTGAGACCTAGCCCCACAGATGCTTCCGAACTCCAGAGACTAATGCGCGCTGCAAG 142
QY 158 ACGTGCGGAGCTGGCTGGCGGACGAGTCACTTCGTAAGAAACAGCGTGA 217
DB 143 ACGTGAAGAGCTCTTCCAGACAGAGTCAAGAGATCCTTCTGAAAGATACGTTGA 202
QY 218 TCGAGTGTACGCGTGC-----GCGCGACGCGC 247
DB 203 TCGAATGTACCGTGTGCGAATGCAGCCGACGACCCCGCTACTAGTCGCGACGCGC 262
QY 248 ACCCGAAACCGGACGCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 307
DB 263 ACCCGAAACCGGACGCGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 322
QY 308 AAGGTACCGGATCATGACAAAGATGAGTGTGAGCGCGCGGACGAAATTCATATGATC 367
DB 323 AAGGTACCGGATCATGACAAAGATGAGTGTGAGCGCGCGGACGAAATTCATATGATC 382
QY 368 TCGAG 372
DB 383 TCGAG 387
RESULT 7
AAZ50493 ID AAZ50493 standard; DNA; 387 BP.
XX AC AAZ50493;
XX DT 23-MAY-2000 (first entry)
XX DE KDEL receptor inhibitor-2 DNA.
XX KDEL receptor inhibitor; heat shock protein; immune response;
XX oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease; ss.
XX OS Chimeric - Mus sp.
XX OS Chimeric - Camelus sp.
XX OS Chimeric - Rattus sp.
XX FH Key Location/Qualifiers
XX FT CDS 10..357
XX FT /tag= a
XX FT /product= "KDEL receptor inhibitor protein"
XX FT sig_peptide 10..68
XX FT /tag= b
XX FT /note= "Derived from mouse Bip"
XX FT mat_peptide 69..354
XX FT /tag= c
XX FT /product= "Mature KDEL receptor inhibitor protein"
XX FT misc_feature 98..222
XX FT /tag= d
XX FT /note= "Rat COMP domain"
XX FT misc_feature 253..324
XX FT /tag= e
XX FT /note= "Camel IgG linker domain"
XX PN WO200006729-A1.
XX PD 10-FEB-2000.
XX PF 28-JUL-1999; 99WO-US17147.
XX PR 29-JUL-1998; 98US-0124671.
XX PA (SLOK) SLOAN KETTERING INST CANCER RES.
XX PI Rothman JE, Mayhew M, Hoe MH;
XX

DR WPI; 2000-195296/17.
DR P-PSDB; AAY44959.
XX
PT Inhibitors of the KDEL receptor which comprises an oligomerization
PT domain useful for promoting secretion of proteins which are normally
PT retained within the cell -
XX
PS Disclosure; Fig 2; 87pp; English.
XX
XX The patent discloses the use of KDEL receptor inhibitor to promote
CC secretion of proteins that are normally retained within the cell such as
CC heat shock proteins by inhibiting KDEL receptor-mediated return of
CC protein complexes to endoplasmic reticulum. This makes the secreted heat
CC shock proteins more accessible to the immune system and improves immune
CC response to a target antigen. The inhibitor protein comprises several
CC subunits where each subunit comprises an oligomerization domain and has
CC at its carboxy terminus a region which binds to a KDEL receptor. The
CC target antigen may be associated with diseases including neoplasia such
CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
CC astrocytoma, with defective tumour suppressor genes, oncogenes,
CC infectious diseases, allergy or autoimmune diseases. The present
CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
CC cleavable signal peptide; the oligomerization domain from rat cartilage
CC oligomeric matrix protein (COMP); a camel IgG linker domain and the
CC carboxy-terminal sequence KDEL. The subsequence GDC is an alteration of
CC rat COMP which provides increased stability via disulphide bonds. This
CC is introduced into host cells by suitable vectors.
XX
SQ Sequence 387 BP; 93 A; 107 C; 126 G; 61 T; 0 other;
Query Match 51.6%; Score 192; DB 21; Length 387;
Best Local Similarity 76.4%; Pred. No. 1.4e-36;
Matches 307; Conservative 0; Mismatches 50; Indels 45; Gaps 4;
QY 1 AAGCTTACATGAGGAGTACATGATTTTAAAGCTTCTGCGCTGGCGGACGTGCAGC 60
DB 1 AAGCTTACATGAGGAGTACATGATTTTAAAGCTTCTGCGCTGGCGGACGTGCAGC 59
QY 61 GCTGCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
DB 60 GCGGCG--CGAGGATTCAGGCTGGGTGGAGCTGTTGT-----CCACAGATG 105
QY 121 CTTGCGAACTGCGAGAAACCAACCGCGCTGCGAGACGTGCGGACTGGCTGGCGAG 180
DB 106 CTTGCGAACTGCGAGAAACCAACCGCGCTGCGAGACGTGCGGACTGGCTGGCGAG 165
QY 181 CAGGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 235
DB 166 CAGGTCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 225
QY 236 -----GCGCGACGCGGACCGGAAACCGCAGCGCGAGCGG 270
DB 226 CAGCGCGACGCGGACCGCGGATCTAGTGTGCGGACCGGAAACCGCAGCGCGAGCGG 285
QY 271 CAGCGCGACGCGGAAACCGGACCGGAAACCGGAAAGTACCGGATCATGAGAA 330
DB 286 CAGCGCGACGCGGAAACCGGACCGGAAACCGGAAAGTACCGGATCATGAGAA 345
QY 331 GATGAGTGTGAGGCGCGGAGAAATTCATATGATGATGATGATGATGATGATGATGATG 372
DB 346 GATGAGTGTGAGGCGCGGAGAAATTCATATGATGATGATGATGATGATGATGATGATG 387
RESULT 8
AAZ50501 ID AAZ50501 standard; DNA; 444 BP.
XX AC AAZ50501;
XX DT 23-MAY-2000 (first entry)
XX DE KDEL receptor inhibitor-10 DNA.
XX

KM KDEL receptor inhibitor; heat shock protein; immune response;
 KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KM infectious disease; allergy; autoimmune disease; ss.
 OS Chimeric - Mus sp.
 OS Chimeric - Rattus sp.
 OS Chimeric - Camelus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 10..414
 FT /tag= a
 FT /product= "KDEL receptor inhibitor protein"
 FT sig_peptide 10..69
 FT /tag= b
 FT /note= "Derived from mouse Bip"
 FT mat_peptide 70..411
 FT /tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"
 FT misc_feature 154..279
 FT /tag= d
 FT /note= "Rat COMP pentamerisation domain"
 FT misc_feature 310..381
 FT /tag= e
 FT /note= "Camel IgG linker domain"
 XX
 PN WO200006729-A1.
 XX
 PD 10-FEB-2000.
 XX
 PF 28-JUL-1999; 99WO-US17147.
 XX
 PR 29-JUL-1998; 98US-0124671.
 XX
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 PI Rothman JE, Mayhew M, Hoe MH;
 XX
 DR MPI: 2000-195296/17.
 P-PSDB; AAY44967.
 XX
 PT Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX
 PS Disclosure; Fig 10; 87bp; English.
 XX
 CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; a myc-tag; an N-glycosylation sequence; the
 CC oligomerisation domain of rat cartilage oligomeric matrix protein
 CC (COMP); a camel IgG linker domain and the carboxy-terminal sequence
 CC KDEL. The subsequence G9CC is an alteration of rat COMP which provides
 CC increased stability via disulphide bonds.
 CC
 SQ Sequence 444 BP; 115 A; 122 C; 134 G; 73 T; 0 other;
 Query Match 51.3%; Score 190.8; DB 21; Length 444;
 Best Local Similarity 80.6%; Pred. No. 2.8e-36;
 Matches 266; Conservative 0; Mismatches 22; Indels 42; Gaps 2;
 73 GGATCCAGCTGGGTGAGACTGTTGTCAGACTGGCCGACAGATGCTTCGGAACTG 132

Db 127 GGATCCAGCTGGGTGAGACTGTTG-----CCACAGATGCTTCGAACTC 174
 QY 133 CAGAAACCAACGCGGCGCTGCAGAGCTGGGAGCTGCGGACGAGTCAAGGAG 192
 Db 175 CAGAGACTAATGCGGCGCTGCAGAGCTGTTGCGACAGCGTCAAGGAG 234
 QY 193 ATCAGTTCTCGAATAACAGCGGTATGAGTGAACGCGTGC----- 235
 Db 235 ATCAGTTCTCGAATAACAGCGGTATGAGTGAACGCGTGC----- 235
 QY 236 -----GCGCCAGCCCGCAGCCGAAACCGGACCGCAGCCGCGCAGCCG 282
 Db 295 ACCCGGTAAGTACGCGACGCGCAGGAAACCGGACCGCAGCCGCGCAGCCG 354
 QY 283 AAACCGCAGCCGAAACCGGAAACCGGAAAGTTCGCGATCATAGAAAAAGATGTTAG 342
 Db 355 AAACCGCAGCCGAAACCGGAAACCGGAAAGTTCGCGATCATAGAAAAAGATGTTAG 414
 QY 343 GCGGCGCGAGAAATTCATATGATCATCTCGAG 372
 Db 415 GCGGCGCGAGAAATTCATATGATCATCTCGAG 444
 RESULT 9
 AA250494
 ID AA250494 standard; DNA; 357 BP.
 XX
 AC AA250494;
 XX
 DT 23-MAY-2000 (first entry)
 XX
 DE KDEL receptor inhibitor-3 DNA.
 XX
 KM KDEL receptor inhibitor; heat shock protein; immune response;
 KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KM infectious disease; allergy; autoimmune disease; ss.
 OS Chimeric - Mus sp.
 OS Chimeric - Camelus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 10..327
 FT /tag= a
 FT /product= "KDEL receptor inhibitor protein"
 FT sig_peptide 10..68
 FT /tag= b
 FT /note= "Derived from mouse Bip"
 FT mat_peptide 69..324
 FT /tag= c
 FT /product= "Mature KDEL receptor inhibitor protein"
 FT misc_feature 97..222
 FT /tag= d
 FT /note= "Mouse TSP3 domain"
 FT misc_feature 223..294
 FT /tag= e
 FT /note= "Camel IgG linker domain"
 XX
 PN WO200006729-A1.
 XX
 PD 10-FEB-2000.
 XX
 PF 28-JUL-1999; 99WO-US17147.
 XX
 PR 29-JUL-1998; 98US-0124671.
 XX
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 XX
 PI Rothman JE, Mayhew M, Hoe MH;
 XX
 DR MPI: 2000-195296/17.
 P-PSDB; AAY44960.

XX Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX
 PS Disclosure; Fig 3, 87pp; English.
 CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerization domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of mouse
 CC thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GPCC is an
 CC alteration of rat cartilage oligomeric matrix protein which provides
 CC increased stability via disulphide bonds.
 XX
 SQ Sequence 357 BP; 84 A; 100 C; 114 G; 59 T; 0 other;
 XX
 Query Match 47.7%; Score 177.6; DB 21; Length 357;
 Best Local Similarity 73.4%; Pred. No. 3.8e-33;
 Matches 2/3; Conservative 0; Mismatches 84; Indels 15; Gaps 3;
 XX
 QY 1 AAGCTTACCATTGGAAGGATGATGATTTAGGCTTGCCCTTGCGGAGCTTCAGC 60
 DB 1 AAGCTTACCATTGGAAGGATGATGATTTAGGCTTGCCCTTGCGGAGCTTCAGC 59
 QY 61 GCTGCCAAAAGATGCCAGCTGGGTGAGACTGTTTTCAGACCTGGGCCCGCAGATG 120
 DB 60 GCGGCGC--CGAGGAGATCCAGCTGGGTGAGACTGTTTTCAGACCTGGGCCCGCAG 114
 QY 121 CTTCGGGAACCTCAGGAACCAAGCGGCGCTGCGAGAGCTGGGAGCTGGCGGCGAG 180
 DB 115 -----CTACCTCTTTCAACCAAGATCTTCTGAGACTTGGGAGCAATCCGAGC 165
 QY 181 CAGGTCAAGGAGATCACTTCTGAAAAACAGGTGATGAGTGCACCGTGCAGGCGC 240
 DB 166 CAGGTGAAGGAATGCTCACTCAACCGAATCATCATGAGTGTGAGTGTGCGGTCCG 225
 QY 241 CAGCCGCAAGCCGAAACCGGACCGCGCAGCCGACCGAACCAGCCGAAACCG 300
 DB 226 CAGCCGCAAGCCGAAACCGGACCGCGCAGCCGACCGAACCAGCCGAAACCG 285
 QY 301 GAACCGGAAGTACCGGATCATGAAAAAGATGAGTTAGGCGGCGCAGAAATTCAT 360
 DB 286 GAACCGGAAGTACCGGATCATGAAAAAGATGAGTTAGGCGGCGCAGAAATTCAT 345
 QY 361 ATGCATCTCGAG 372
 DB 346 ATGCATCTCGAG 357
 XX
 RESULT 10
 ID AAA47734 standard; DNA; 755 BP.
 XX AAA47734;
 AC
 XX 08-NOV-2000 (first entry)
 DE Human COMP/TSP-1 chimeric protein coding sequence.
 XX
 XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;
 XX

KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glioma; ds.
 XX
 XX Homo sapiens.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT CDS 26..754
 FT /tag= a
 FT /product= COMP/TSP-1 chimeric protein
 XX
 PN MO200044908-A2.
 XX
 XX 03-AUG-2000.
 PD
 XX 01-FEB-2000; 2000MO-US02482.
 PF
 XX 01-FEB-1999; 99US-0118053.
 PR
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA
 XX Lawler JW;
 PI
 XX WPI: 2000-514823/46.
 DR
 XX P-PSDB; AAB00040.
 XX
 PT Nucleic acids encoding chimeric proteins such as cartilage oligomeric
 PT matrix protein (COMP)/thrombospondin (TSP)-1 and 2, useful for
 PT inhibiting angiogenesis and treating diseases such as cancer
 XX
 PS Claim 11; Fig 4a-b; 40pp; English.
 XX
 CC New nucleic acids are described which encode a protein comprising
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
 CC but not the TSP (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating
 CC angiogenesis related diseases such as cancer (by reducing the rate of
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 XX
 SQ Sequence 755 BP; 147 A; 245 C; 236 G; 127 T; 0 other;
 XX
 Query Match 38.3%; Score 142.6; DB 21; Length 755;
 Best Local Similarity 97.3%; Pred. No. 9.8e-25;
 Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 XX
 QY 98 GTTCAGACCTGGGCGCCGAGATGCTTTCGGAAGTGCAGAAACCAACCGCGCTGCAG 157
 DB 105 GCTCAGACTGGGCGCCGAGATGCTTTCGGAAGTGCAGAAACCAACCGCGCTGCAG 164
 QY 158 ACGTCCGGAGCTGCTGCGGCAAGTCAAGAGATCACTTCTGAAAAACACGGTGA 217
 DB 165 ACGTCCGGAGCTGCTGCGGCAAGTCAAGAGATCACTTCTGAAAAACACGGTGA 224
 QY 218 TGGAGTGTGACGCGTGGCGGCGCGCAGCCG 246
 DB 225 TGGAGTGTGACGCGTGGCGGATGCAGCAG 253
 XX
 RESULT 11
 ID AAA47735 standard; DNA; 925 BP.
 XX AAA47735;
 XX

AC	AAA47735;	
XX		
DT	08-NOV-2000	(first entry)
XX		
DE	Human COMP/TSP-2 chimeric protein coding sequence.	
XX		
KW	TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;	
KW	thrombospondin; angiogenesis; tumour; treatment; cancer;	
KW	arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;	
KW	glaucoma; ds.	
XX		
OS	Homo sapiens.	
XX		
OS	Synthetic.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	26..925
FT		/tag= a
FT		/product= COMP/TSP-2 chimeric protein
XX		
PN	WO200044908-A2.	
XX		
PD	03-AUG-2000.	
XX		
PF	01-FEB-2000; 2000MO-US02482.	
XX		
PR	01-FEB-1999; 99US-0118053.	
XX		
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.	
XX		
PI	Lawler JW;	
XX		
DR	WPI; 2000-514823/46.	
DR	P-PSDB; AAB00041.	
XX		
PT	Nucleic acids encoding chimeric proteins such as cartilage oligomeric	
PT	matrix protein (COMP)/thrombospondine (TSP)-1 and 2, useful for	
PT	inhibiting angiogenesis and treating diseases such as cancer	
XX		
PS	Claim 44; Fig 5a-b; 40pp; English.	
XX		
CC	New nucleic acids are described which encode a protein comprising	
CC	the second and third type 1 repeats of human TSP (thrombospondin)-1,	
CC	but not the TGF (transforming growth factor)-beta activation region.	
CC	of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing	
CC	the second and third type-1 repeats and the COMP (cartilage	
CC	oligomeric matrix protein) assembly sequence (COMP/TSP-1) was	
CC	produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1	
CC	caused inhibition of the growth of tumours in mice models.	
CC	Thus the nucleic acids and proteins may be useful for treating	
CC	angiogenesis related diseases such as cancer (by reducing the rate of	
CC	growth and size of tumours), arthritis, psoriasis, diabetic	
CC	retinopathy, corneal graft rejection, and glaucoma. They may also be	
CC	used for treating human immunodeficiency virus (HIV) infection.	
CC	Anti-angiogenic therapy has little toxicity, does not require the	
CC	therapeutic agent to enter tumour cells or cross the blood-brain	
CC	barrier, controls tumour growth independently of growth of	
CC	tumour cell heterogeneity, and does not induce drug resistance.	
XX		
XX		
SQ	Sequence 925 BP; 156 A; 310 C; 312 G; 147 T; 0 other;	
XX		
QY	Query Match	38.3%; Score 142.6; DB 21; Length 925;
DB	Best Local Similarity	97.3%; Pred. No. 1e-24;
DB	Matches 145; Conservative	0; Mismatches 4; Indels 0; Gaps 0
QY	98 GTTCAGACTGGGCGCCGAGATGCTTCGGGAATCAGAAACCAACGGCGCTGCAG	157
DB	105 GCTCAGACTGGGCGCCGAGATGCTTCGGGAATCAGAAACCAACGGCGCTGCAG	164
QY	158 ACGTCGGGAGCTGCGTCGGCGCAGAGTTCAGGAGATCAGTCTCTGAAACACGGTGA	217
DB	165 ACGTCGGGAGCTGCGTCGGCGCAGAGTTCAGGAGATCAGTCTCTGAAACACGGTGA	224
QY	218 TGAAGTGAACCGCTGGCGGCGCCGACGCC	246

Query Match	38.3%	Score 142.6	DB 24	Length 2439
Best Local Similarity	97.3%	Pred. No. 1.2e-24		
Matches 145	Conservative 0	Mismatches 4	Indels 0	Gaps 0
98	GTTCAGACCTGGGGCCGCGAGATGCTTGGGAACTCGAGAAACCAACGCGCGCTGAGG	157		
105	GCTAGACCTGGGGCCGCGAGATGCTTGGGAACTCGAGAAACCAACGCGCGCTGAGG	164		
158	ACGTGCGGGAATGCGCTCGCGAGAGGTCAAGGATCAAGTTCCTGAAAAACAAGCTGA	217		
165	ACGTGCGGGAATGCGCTCGCGAGAGGTCAAGGATCAAGTTCCTGAAAAACAAGCTGA	224		
218	TGAGGTGTGAAGCTGTCGGGCGCGAGCCG	246		

XX (AVAL-) AVALON PHARM.
PA Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 441; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
SQ Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;
XX
Query Match 38.3%; Score 142.6; DB 24; Length 2439;
Best Local Similarity 97.3%; Pred. No. 1.2e-24;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY 98 GTTCGACCTGGGCGCGGAGAGTCTCGGGAACCTGAGAAACCAACGGCGGTGCAGG 157
DB 105 GCTCAGACCTGGGCGCGGAGAGTCTCGGGAACCTGAGAAACCAACGGCGGTGCAGG 164
QY 158 ACGTCGGGAGCTGCTGCGCGGAGAGTCTCGGGAACCTGAGAAACCAACGGGTGA 217
DB 165 ACGTCGGGAGCTGCTGCGCGGAGAGTCTCGGGAACCTGAGAAACCAACGGGTGA 224
QY 218 TGGAGTGTGACCGGTGCGGCGCGGAGCGG 246
DB 225 TGGAGTGTGACCGGTGCGGCGGAGTGCAGAG 253
XX
RESULT 15
ABL62870
ID ABL62870 standard; DNA; 2439 BP.
XX
AC ABL62870;
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:1207.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
KW cytostatic; gene therapy; anti-neoplastic; Wilms' tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
XX MO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US10838.

XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237298P.
PR 02-OCT-2000; 2000US-237299P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 03-OCT-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 1207; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine

CC carcinoma, papillary carcinoma and wilm's tumour.

XX Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;

Query Match 38.3%; Score 142.6; DB 24; Length 2439;

Best Local Similarity 97.3%; Pred. No. 1.2e-24;

Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 98 GTTCAGACCTGGGGCCCGCAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 157

Db 105 GCTCAGACCTGGGGCCCGCAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 164

Qy 158 ACGTGCGGACTGCTGCGGCGCAGCAGGTCAAGGATCACTTCTGAAAAACACGGTGA 217

Db 165 ACGTGGGGACTGCTGCGGCGCAGCAGGTCAAGGATCACTTCTGAAAAACACGGTGA 224

Qy 218 TGGAGGTGACCGCTGCGGGCCCGCAGCCG 246

Db 225 TGGAGGTGACCGCTGCGGGCGATGCAGCAG 253

Search completed: August 26, 2003, 16:08:06
Job time : 209 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:10:25 ; Search time 55 Seconds
(without alignment)
2985.355 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372

Sequence: 1 aagcttaccatggaagta.....attcatatgcattcgag 372

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 113956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	ID	Description
	1	372	100.0	372	3	US-09-124-671-24	Sequence 24, Appl
	2	271.2	72.9	372	3	US-09-124-671-30	Sequence 30, Appl
	3	264.8	71.2	372	3	US-09-124-671-28	Sequence 28, Appl
	4	202	54.3	369	3	US-09-124-671-22	Sequence 22, Appl
	5	194	52.2	369	3	US-09-124-671-20	Sequence 20, Appl
	6	193.4	52.0	387	3	US-09-124-671-14	Sequence 14, Appl
	7	192	51.6	387	3	US-09-124-671-16	Sequence 16, Appl
	8	190.8	51.3	444	3	US-09-124-671-35	Sequence 35, Appl
	9	177.6	47.7	357	3	US-09-124-671-18	Sequence 18, Appl
	10	156.4	42.0	315	3	US-09-124-671-26	Sequence 26, Appl
	11	56.4	15.2	480	5	PCT-US96-04648-1	Sequence 1, Appl
	12	55.4	14.9	397	3	US-09-253-691-3	Sequence 3, Appl
	13	55	14.8	1037	4	US-09-181-585-3	Sequence 3, Appl
	14	55	14.8	1159	4	US-09-181-585-1	Sequence 1, Appl
	15	55	14.8	1471	4	US-09-181-585-2	Sequence 1, Appl
	16	53	14.2	521	4	US-09-643-597-334	Sequence 324, App
	17	53	14.2	521	4	US-09-480-864A-324	Sequence 324, App
	18	53	14.2	521	4	US-09-542-615A-324	Sequence 324, App
	19	53	14.2	521	4	US-09-606-421B-324	Sequence 324, App
	20	52.6	14.1	477	4	US-09-135-994-1	Sequence 1, Appl
	21	52.6	14.1	477	4	US-09-684-843A-1	Sequence 1, Appl
	22	52.2	14.0	23187	4	US-09-499-523-1	Sequence 1, Appl
	23	51.4	13.8	234	1	US-08-469-802B-3	Sequence 3, Appl
	24	51.4	13.8	234	1	US-08-267-803B-3	Sequence 3, Appl
	25	50.4	13.5	71	1	US-08-464-318-11	Sequence 11, Appl
	26	50.4	13.5	71	2	US-08-471-341-11	Sequence 11, Appl
	27	50.4	13.5	71	2	US-08-461-566-11	Sequence 11, Appl

28	49.8	13.4	336	2	US-07-814-220-4	Sequence 4, Appl
29	49.8	13.4	336	2	US-07-812-421-4	Sequence 4, Appl
30	49.6	13.3	10348	2	US-08-457-273B-41	Sequence 41, Appl
31	49.6	13.3	10348	3	US-08-556-419-13	Sequence 13, Appl
32	49.6	13.3	10348	3	US-09-041-886-14	Sequence 14, Appl
33	49.6	13.3	10366	1	US-08-246-982A-5	Sequence 5, Appl
34	49.6	13.3	10366	1	US-08-453-265-5	Sequence 5, Appl
35	49.2	13.2	203	3	US-09-043-303-7	Sequence 7, Appl
36	48.4	13.0	71	1	US-08-464-318-12	Sequence 12, Appl
37	48.4	13.0	71	2	US-08-471-341-12	Sequence 12, Appl
38	48.4	13.0	71	2	US-08-461-566-12	Sequence 12, Appl
39	48.2	13.0	336	2	US-07-814-220-3	Sequence 3, Appl
40	48.2	13.0	336	2	US-07-812-421-3	Sequence 3, Appl
41	47.8	12.8	90	1	US-08-464-318-8	Sequence 8, Appl
42	47.8	12.8	90	2	US-08-471-341-8	Sequence 8, Appl
43	47.8	12.8	90	2	US-08-461-566-8	Sequence 8, Appl
44	47.4	12.7	2294	4	US-09-086-663A-70	Sequence 70, Appl
45	47.4	12.7	3334	4	US-09-086-663A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-124-671-24
Sequence 24, Application US/09124671A
Patent No. 6160088
GENERAL INFORMATION:
APPLICANT: Rothman, James
APPLICANT: Mayhew, Mark
TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
FILE REFERENCE: 31488
CURRENT APPLICATION NUMBER: US/09/124,671A
CURRENT FILING DATE: 1998-07-29
NUMBER OF SEQ ID NOS: 42
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 24
LENGTH: 372
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chimeric human COMP-KDEL
US-09-124-671-24

DP?

Query Match	100.0%; Score 372; DB 3; Length 372;
Best Local Similarity	100.0%; Pred. No. 3e-81;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 AAGCTTACCATGGGAAGGATCATATTTAGCTTGGCTGCGCCCTTGGGCGAGTCGACG 60
DB	1 AAGCTTACCATGGGAAGGATCATATTTAGCTTGGCTGCGCCCTTGGGCGAGTCGACG 60
QY	61 GCTGCCAAAAAAGATCCAGCTGGTGAGACTGTTGTCAGACTTGGGCGCCGACATG 120
DB	61 GCTGCCAAAAAAGATCCAGCTGGTGAGACTGTTGTCAGACTTGGGCGCCGACATG 120
QY	121 CTTGGGAATGTCAGAAACCAACGCGCGCTGACAGACTGCGGAGCTGCTCGCGAG 180
DB	121 CTTGGGAATGTCAGAAACCAACGCGCGCTGACAGACTGCGGAGCTGCTCGCGAG 180
QY	181 CAGTCCAGGAGATCAGCTTCTGTAACCAACGCTGATGATGAGTGCAGCTGCGGGCGG 240
DB	181 CAGTCCAGGAGATCAGCTTCTGTAACCAACGCTGATGATGAGTGCAGCTGCGGGCGG 240
QY	241 CAGCCGACGCGAAGCCGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB	241 CAGCCGACGCGAAGCCGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY	301 GAACCGGAAGGTACCGATCATCAGAAAAGATGATGTTGAGCGGCGGCGGCGGCGGCG 360
DB	301 GAACCGGAAGGTACCGATCATCAGAAAAGATGATGTTGAGCGGCGGCGGCGGCGGCG 360

Db	128	TAACCCAAATGAAATCAGATGCTGGGAGAGCTCCGAGATGTGATGAGACAGCAGGTTGAAAG	187
Qy	191	AGATCAAGTTCCTGAAAAACACGGTGAATGAGATGAGACGCTGGGGCCGACGCGCAGC	250
Db	188	AGACCATGTTCTTGAGAAACACCATTCGAGAAATGCCAGGCTCTGTGGCCCGACGCCGAGC	247
Qy	251	CGAAATCCGACGCCGACGCCGACGCCGACGCCGAAACCGCACCCGAAACCGGAAACCGGAAAG	310
Db	248	CGAAATCCGACGCCGACGCCGACGCCGACGCCGAAACCGCACCCGAAACCGGAAACCGGAAAG	307
Qy	311	GTACCGGATCATCAGAAAAAGATGATTGTAGAGCGCGCAGAAATTCATATGATCTCG	370
Db	308	GTACCGGATCATCAGAAAAAGATGATTGTAGAGCGCGCGCAGAAATTCATATGATCTCG	367
Qy	371	AG 372	
Db	368	AG 369	

```

RESULT 5
US-09-124-671-20
/ Sequence 20, Application US/09124671A
/ Patent No. 6,60098
/ GENERAL INFORMATION:
/ APPLICANT: Rothman, James
/ APPLICANT: Maynew, Mark
/ APPLICANT: Hoe, Mee
/ TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
/ FILE REFERENCE: 31488
/ CURRENT APPLICATION NUMBER: US/09/124,671A
/ CURRENT FILING DATE: 1998-07-29
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 20
/ LENGTH: 369
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURES:
/ OTHER INFORMATION: chimeric mouse TSP1-KDEL
/ US-09-124-671-20

```

Query Match	52.2%	Score 194	DB 3	Length 369
Best Local Similarity	77.5%	Pred. No. 2.6e-38		
Matches 234	Conservative	0	Mismatches 68	Indels 0
			Gaps	0

Qy	71	AAGGATCTCAGCCTGGGTGAGACTTGTTCAGACTTGCGGCCCAATGCTTGGGAAAC	130
Db	68	AAGGATCTCAGCCTGGGTGAGACTTGTTCAGACTTGCGGGAACCAAGGACTGGTCAACCCAGC	127
Qy	131	TGCAGGAAACCAACGGCGGCGCTTCAGAGAGTGTCGGGAATGCTGCTGGGAGCAGGTCAAGG	190
Db	128	TCACCTCTTCAACCGAGATCTAGTGAGACTTCGGGAGAGACATCCGAGACCGGTGAAAG	187
Qy	191	AGATCACTGTCTCTGAAAAAACGAGTGATGAGATGGAACGCTGCGGGGCGGAGCCGAGC	250
Db	188	AATGTCACTACTCCGGAAACCACTCATATGAGATGTCAATGTGTGGGTGCCAGCCGAGC	247
Qy	251	CGAAATCCGAGCGCGCAGCCGCAAGCCGCAACCCGAGCCGAAACCGGAAACCGGAAAG	310
Db	248	CGAAACCGGAGCGCGCAGCCGCAAGCCGCAACCCGAAACCGGAAACCGGAAACCGGAAAG	307
Qy	311	GTCACCGATCATCAGAAAAAGATGAGTTGATAGCGCGCGCAGAAATTCATATGATCTCG	370
Db	308	GTCACCGATCATCAGAAAAAGATGAGTTGATAGCGCGCGCAGAAATTCATATGATCTCG	367
Qy	371	AG 372	
Db	368	AG 369	

RESULT 6
US-09-124-671-14

Sequence 14, Application US/09124671A
 Patent No. 6160088
 GENERAL INFORMATION:
 APPLICANT: Rothman, James
 APPLICANT: Mayhew, Mark
 APPLICANT: Hoe, Mee
 TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
 FILE REFERENCE: 31488
 CURRENT APPLICATION NUMBER: US/09/124,671A
 CURRENT FILING DATE: 1998-07-29
 NUMBER OF SEQ ID NOS: 42
 SOFTWARE: FASTSEQ For Windows Version 3.0
 SEQ ID NO 14
 LENGTH: 387
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: chimeric rat COMP-KDEL
 US-09-124-671-14

Query Match	52.0%	Score 193.4	DB 3	Length 387
Best Local Similarity	81.6%	Pred. No. 3.6e-38		
Matches 249	Conservative 0	Mismatches 26	Indels 30	Gaps 1

Db	83	GTGGAGACCTGATGCCCAAGATGCTTCCAGAACTTCAAGAACTAATGCGGCGCTGCAG	1 42
Qy	158	ACCTTGGGGGACTGGCTGCGCGCAGCAGGTCAAGGAGATCACTTCTGAAAAACAGCGTGA	21 7
Db	143	ACGTGAGAGAGCTCTTGGCCACAGCAGGTCAAGAGATCACTTCTGAAAGATACGGTGA	202
Qy	218	TGGAGTGTACCGGTCG-----GGCGCAGCGCG	24 7
Db	203	TGGAATGTGACGTTGTCCGAATGCAGCCCGCAGCACCCCGGTACTAAGTCCGAGCGCG	263
Qy	248	AGCCGAAACCGCAGCGCGCAGCGCAGCGCGAAACCGCAGCCGAAACCGGAACCGG	307
Db	263	AGCCGAAACCGCAGCGCGCAGCGCAGCGCGAAACCGCAGCCGAAACCGGAACCGG	322
Qy	308	AAGGTACCGGATCATATAGAAAAAGTAGTTGTAGGCGGCGCGCAAAATTCATATGCATC	367
Db	323	AAGGTACCGGATCATATAGAAAAAGTAGTTGTAGGCGGCGCGCAAAATTCATATGCATC	382
Qy	368	TCGAG	372
Db	383	TCGAG	387

RESULT 7
US-09-124-671-16
; Sequence 16, Application US/09124671A

```

: TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
:
: FILE REFERENCE: 31488
:
: CURRENT APPLICATION NUMBER: US/09/124,671A
:
: CURRENT FILING DATE: 1998-07-29
:
: NUMBER OF SEQ ID NOS: 42
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ ID NO: 16
:
: LENGTH: 387
:
: TYPE: DNA
:
: ORGANISM: Artificial Sequence
:
: FEATURE:
:
: OTHER INFORMATION: chimeric rat COMP-KDEL
:
: US-09-124-671-16

```

Query Match	51.6%;	Score 192;	DB 3;	Length 387;
Best Local Similarity	76.4%;	Pred. No. 7.8e-38;		

Matches 307; Conservative 0; Mismatches 50; Indels 45; Gaps 4;

QY 1 AAGCTTACCATGGAAGATGATGATTTAGGCTTGCCTGCTGCGGAGTCTGCAGC 60
Db 1 AAGCTTACCATGGAAGATGATGATGATTTAGGCTTGCCTGCTGCGGAGTCTGCAGC 59
QY 61 GCTGCCAAAAAGATCCAGCTGGGTGAGAGTGTGTTGACAGCTTGAGCCCGCAGATG 120
Db 60 GCGGGC--CGAGGGATCCAGCTGGGTGAGAGTGTGTTGATGATGATGATGATGATG 105
QY 121 CTTCCGGAAGTCCAGGAAACCAACGCGGCTGCGAGAGTGGGAGCTGGGCGGCGAG 180
Db 106 CTTCCGGAAGTCCAGGAAACCAACGCGGCTGCGAGAGTGGGAGCTGGGCGGCGAG 165
QY 181 CAGGTCAGGAGATCACTGTTCTGAAAAA CACGGTGTGAGTGTGATGATGATGATG 235
Db 166 CAGGTCAGGAGATCACTGTTCTGAAAAA CACGGTGTGAGTGTGATGATGATGATG 225
QY 236 -----GCGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 270
Db 226 CAGCGCGACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 285
QY 271 CAGCGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 330
Db 286 CAGCGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 345
QY 331 GATGAGTTGTAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 372
Db 346 GATGAGTTGTAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 387

RESULT 8
US-09-124-671-35
; Sequence 35, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: KDEL-myc
US-09-124-671-35

Query Match 51.3%; Score 190.8; DB 3; Length 444;
Best Local Similarity 80.6%; Pred. No. 1.6e-37;
Matches 266; Conservative 0; Mismatches 22; Indels 42; Gaps 2;

QY 73 GATCCAGGCTGGTGGAGAGTGTGTCAGACTGGGCGCGCAGATGCTTGGGAACTG 132
Db 127 GGATCCAGGCTGGTGGAGAGTGTGTCAGACTGGGCGCGCAGATGCTTGGGAACTG 174
QY 133 CAGGAAACCAACGCGGCTGCGAGAGTGGGCGGAGTGGCTGCGGAGAGTCAAGGAG 192
Db 175 CAGGAGACTAAATGCGGCGCTGCAAGAGTGAAGAGAGCTTTCGACAGAGGTCAGAG 234
QY 193 ATCAGCTTCTGAAAAACACGCGTGTGAGTGTGAGCGCTGGG----- 235
Db 235 ATCAGCTTCTGAAAAACACGCGTGTGAGTGTGAGCGCTGGG----- 294
QY 236 -----GCGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 282
Db 295 ACCCGCGGACTAGTCCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 354

QY 283 AAACCGGACCGGAAACCGGAAACCGGAAAGTACCGGATCATGAAAAAGATGATTGAG 342
Db 355 AAACCGGACCGGAAACCGGAAACCGGAAAGTACCGGATCATGAAAAAGATGATTGAG 414
QY 343 GCGGCGGAGAAATTCATATGATGATCTCGAG 372
Db 415 GCGGCGGAGAAATTCATATGATGATCTCGAG 444

RESULT 9
US-09-124-671-18
; Sequence 18, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 357
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric mouse TSP1-KDEL
US-09-124-671-18

Query Match 47.7%; Score 177.6; DB 3; Length 357;
Best Local Similarity 73.4%; Pred. No. 2.3e-34;
Matches 273; Conservative 0; Mismatches 84; Indels 15; Gaps 3;

QY 1 AAGCTTACCATGGAAGATGATGATTTAGGCTTGCCTGCTGCGGAGTCTGCAGC 60
Db 1 AAGCTTACCATGGAAGATGATGATTTAGGCTTGCCTGCTGCGGAGTCTGCAGC 59
QY 61 GCTGCCAAAAAGATCCAGCTGGGTGAGAGTGTGTTGACAGCTTGAGCCCGCAGATG 120
Db 60 GCGGGC--CGAGGGATCCAGCTGGGTGAGAGTGTGTTGATGATGATGATGATGATG 114
QY 121 CTTCCGGAAGTCCAGGAAACCAACGCGGCTGCGAGAGTGGGAGTGGCTGGGCGAG 180
Db 115 -----CTCACCTCTTCAACCAAGATCCTAGTGAAGCTTCCGGAACGATCCGAGAC 165
QY 181 CAGGTCAGGAGATCACTGTTCTGAAAAA CACGGTGTGAGTGTGAGCGCTGGGCGG 240
Db 166 CAGGTCAGGAGATCACTGTTCTGAAAAA CACGGTGTGAGTGTGAGCGCTGGGCGG 225
QY 241 CAGCGCGACGCGGAAACCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 226 CAGCGCGACGCGGAAACCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 285
QY 301 GAACCGGAAGTATCCGATCATCAAAAAAGATGATGTTGAGCGGCGGCGGAAATTCAT 360
Db 286 GAACCGGAAGTATCCGATCATCAAAAAAGATGATGTTGAGCGGCGGCGGCGGAAATTCAT 345
QY 361 ATGCACTTCGAG 372
Db 346 ATGCACTTCGAG 357

RESULT 10
US-09-124-671-26
; Sequence 26, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS


```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

RESULT 13
US-09-181-585-3/c
; Sequence 3, Application US/09181585
; Patent No. 6524791
; GENERAL INFORMATION:
; APPLICANT: Rannu, Laura P.W.
; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 11000900101
; CURRENT APPLICATION NUMBER: US/09/181,585

```

/ CURRENT FILING DATE: 1998-10-28
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO: 3
/ LENGTH: 1037
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: cDNA
/ OTHER INFORMATION: comprising exons E, C, and A
US-09-181-585-3

Query Match 14.8%; Score 55; DB 4; Length 1037;
Best Local Similarity 53.5%; Pred. No. 0.0001;
Matches 115; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 94 TGTGTTGAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACCGCGGCTG 153
DB 928 TATTTTAAAAAATGACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 869
QY 154 CAGGACGTGCGGGAAGTGGCTGCGGACGACGACGACGACGACGACGACGACGACGACG 213
DB 868 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 809
QY 214 GTGATGAGTGTGACGCGTGGCGGCGCGACGCCGAAACCGCAGCCGCGCAG 273
DB 808 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 749
QY 274 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGA 308
DB 748 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 714

RESULT 14
US-09-181-585-1/C
/ Sequence 1, Application US/09181585
/ Patent No. 6524791
/ GENERAL INFORMATION:
/ APPLICANT: Rannum, Laura P. W.
/ APPLICANT: Koob, Michael
/ TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
/ FILE REFERENCE: 1100900101
/ CURRENT APPLICATION NUMBER: US/09/181,585
/ CURRENT FILING DATE: 1998-10-28
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO: 1
/ LENGTH: 1159
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-181-585-1

Query Match 14.8%; Score 55; DB 4; Length 1159;
Best Local Similarity 53.5%; Pred. No. 0.0001;
Matches 115; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 94 TGTGTTGAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACCGCGGCTG 153
DB 740 TATTTTAAAAAATGACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 681
QY 154 CAGGACGTGCGGGAAGTGGCTGCGGACGACGACGACGACGACGACGACGACGACGACG 213
DB 680 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 621
QY 214 GTGATGAGTGTGACGCGTGGCGGCGCGACGCCGAAACCGCAGCCGCGCAG 273
DB 620 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 561
QY 274 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGA 308
DB 560 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 526

RESULT 15
US-09-181-585-2/C
/ Sequence 2, Application US/09181585
/ Patent No. 6524791
/ GENERAL INFORMATION:
/ APPLICANT: Rannum, Laura P. W.
/ APPLICANT: Koob, Michael
/ TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
/ FILE REFERENCE: 1100900101
/ CURRENT APPLICATION NUMBER: US/09/181,585
/ CURRENT FILING DATE: 1998-10-28
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO: 2
/ LENGTH: 1471
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: cDNA
/ OTHER INFORMATION: comprising exons D, C, B, and A
US-09-181-585-2

Query Match 14.8%; Score 55; DB 4; Length 1471;
Best Local Similarity 53.5%; Pred. No. 0.00011;
Matches 115; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 94 TGTGTTGAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACCGCGGCTG 153
DB 1362 TATTTTAAAAAATGACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 1303
QY 154 CAGGACGTGCGGGAAGTGGCTGCGGACGACGACGACGACGACGACGACGACGACGACG 213
DB 1302 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1243
QY 214 GTGATGAGTGTGACGCGTGGCGGCGCGACGCCGAAACCGCAGCCGCGCAG 273
DB 1242 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1183
QY 274 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGA 308
DB 1182 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1148

Search completed: August 26, 2003, 15:35:31
Job time : 56 secs

GenCore version 5.1.6
(c) 1993 - 2003 Compu

ic search, using sw model

August 26, 2003, 15:34:34 ; Search time 805 Seconds

- (without alignments)
1038.903 Million cell updates/sec

IS-09-696-872-24

72
aaagcttaccatgggaagta.....aatccatatgcattctcgaag 372

IDENTITY_NUC
apop 10_0 , Gapext 1.0

517243 seqs, 1124081882 residues

its satisfying chosen parameters: 3034486

```
length: 0
mgth: 20000000000
```

Minimum	Match	0%
---------	-------	----

Listing first 45 summaries

Published_Applications_NA: *

```

: /cgn2_6/prodataa/1/pubpna/US07_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/PCT_NEW_PUB.seq.*
: /cgn2_6/prodataa/1/pubpna/US06_NEW_PUB.seq.*
: /cgn2_6/prodataa/1/pubpna/US06_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US07_NEW_PUB.seq.*
: /cgn2_6/prodataa/1/pubpna/PCTUS_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US08_NEW_PUB.seq.*
: /cgn2_6/prodataa/1/pubpna/US08_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US09_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US09B_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US09C_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US09_NEW_PUB.seq.*
: /cgn2_6/prodataa/1/pubpna/US10A_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US10B_PUBCOMB.seq.*
: /cgn2_6/prodataa/1/pubpna/US10_NEW_PUB.seq.*
: /cgn2_6/prodataa/1/pubpna/US60_NEW_PUB.seq.*
: /cgn2_6/prodataa/1/pubpna/US60_PUBCOMB.seq.*

```

is the number of results predicted by chance to have a value greater than or equal to the score of the result being printed, divided by analysis of the total score distribution.

SUMMARIES

Query	Length	DB	ID	Description
38.3	755	10	US-09-919-603-4	Sequence 4, App1
38.3	925	10	US-09-919-603-6	Sequence 6, App1
38.3	2439	10	US-09-954-531-140	Sequence 140, App
38.3	2439	10	US-09-954-531-359	Sequence 359, App
38.3	2439	11	US-09-918-6224B-49	Sequence 49, App1
38.3	2439	12	US-10-101-822-40	Sequence 40, App1
38.3	2439	14	US-10-177-293-71	Sequence 71, App1
38.3	34427	11	US-09-111-911-5	Sequence 5, App1
15.6	35937	10	US-09-782-378A-3	Sequence 3, App1
15.5	1734	14	US-10-156-761-1	Sequence 4994, App
15.5	9025608	14	US-10-156-761-1	Sequence 1, App1
14.5	428	11	US-09-918-995-4559	Sequence 4559, App
14.5	433	11	US-09-918-995-4560	Sequence 4560, App
14.2	521	9	US-09-735-705-324	Sequence 324, App
14.2	521	10	US-09-850-7176A-324	Sequence 324, App
14.2	521	10	US-09-897-778-324	Sequence 324, App

17	53	14.2	521	12	US-10-117-982-324	Sequence 324, Appl
18	53	14.2	1082	9	US-09-925-301-441	Sequence 441, Appl
19	53	14.2	1412	2	US-10-501-822-154	Sequence 154, Appl
20	53	14.2	3747	13	US-10-071-766-33	Sequence 33, Appl
21	52.6	14.1	41936	10	US-09-967-768A-116	Sequence 116, Appl
22	51.2	13.8	96	13	US-10-080-059-14	Sequence 14, Appl
23	51.2	13.8	88421	10	US-09-976-059-11	Sequence 1, Appl
24	50.6	13.6	9025608	14	US-10-156-761-1	Sequence 1, Appl
25	49.6	13.3	614	14	US-10-215-432-30	Sequence 30, Appl
26	49.6	13.3	614	14	US-10-215-432-31	Sequence 31, Appl
27	48.2	13.0	1080	14	US-10-215-432-26	Sequence 26, Appl
28	47.8	12.8	372	13	US-10-079-623-187	Sequence 187, Appl
29	46.4	12.5	1926	12	US-10-294-804-3	Sequence 3, Appl
30	46.4	12.5	2614	11	US-09-882-846-491	Sequence 491, Appl
31	46.4	12.5	6604	10	US-09-820-107-1748	Sequence 1748, Appl
32	46.4	12.5	8705	14	US-10-291-230-14	Sequence 14, Appl
33	46.4	12.5	8705	14	US-10-291-249-14	Sequence 14, Appl
34	46.2	12.4	2276	12	US-10-137-870-9	Sequence 9, Appl
35	46.2	12.4	2276	12	US-10-140-018-9	Sequence 9, Appl
36	46.2	12.4	2276	12	US-10-140-021-9	Sequence 9, Appl
37	46.2	12.4	2276	12	US-10-140-274-9	Sequence 9, Appl
38	46.2	12.4	2276	12	US-10-140-471-9	Sequence 9, Appl
39	46.2	12.4	2276	12	US-10-140-807-9	Sequence 9, Appl
40	46.2	12.4	2276	12	US-10-140-922-9	Sequence 9, Appl
41	46.2	12.4	2276	12	US-10-140-924-9	Sequence 9, Appl
42	46.2	12.4	2276	12	US-10-140-926-9	Sequence 9, Appl
43	46.2	12.4	2276	12	US-10-141-698-9	Sequence 9, Appl
44	46.2	12.4	2276	12	US-10-141-702-9	Sequence 9, Appl
45	46.2	12.4	2276	12	US-10-141-704-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-919-603-4

Patent No. US20020137679A1

APPLICANT: Lawler, John W

; TITLE OF INVENTION: Chitin

CURRENT APPLICATION NUMBER

; PRIOR APPLICATION NUMBER:

; PRIOR APPLICATION NUMBER:

NUMBER OF SEQ ID NOS: 21

; SEQ ID NO 4

TYPE: DNA

FEATURE:

US-09-919-603-4

Query Match	38
-------------	----

Matches 145; Conservativ

QY 98 GTTCAGACCTGGGC

Db 105 GCTCAGACCTGGGC

QY 158 ACGTGGGACTGC

Db 165 ACGTGGGGACTGG

QY 218 TGGAGTGTGACGCC

•

```

Db      225  TGGAGTGTGACGCGCTGCGGAGTGCACGACG  253

RESULT 2
US-09-919-603-6
: Sequence 6, Application US/09919603
: Patent No. US20020137679A1
: GENERAL INFORMATION:
: APPLICANT: Lawler, John W.
: TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
: FILE REFERENCE: 1440.1033-007
: CURRENT APPLICATION NUMBER: US/09/919,603
: PRIOR FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: PCT/US00/02482
: PRIOR FILING DATE: 2000-02-01
: PRIOR APPLICATION NUMBER: 60/118,053
: PRIOR FILING DATE: 1999-02-01
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 925
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: fusion gene
US-09-919-603-6

Query Match      38.3%; Score 142.6; DB 10; Length 925;
Best Local Similarity 97.3%; Pred. No. 3.8e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      98  GTTCAGACCTGGGCGCCGCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG  157
Db      105  GCTCAGACCTGGGCGCCGCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG  164

QY      158  ACGTCCGGAGACTGCGCTGGCGCAGCAGAGTCAAGGAGATACGTTCTGAAAAACCGGTGA  217
Db      165  ACGTCCGGAGACTGCGCTGGCGCAGCAGAGTCAAGGAGATACGTTCTGAAAAACCGGTGA  224

QY      218  TGGAGTGTGACGCGCTGCGGCGCGCGAGCCG  246
Db      225  TGGAGTGTGACGCGCTGCGGCGAGTGCACGACG  253

RESULT 3
US-09-954-531-140
: Sequence 140, Application US/09954531
: Patent No. US20020165180A1
: GENERAL INFORMATION:
: APPLICANT: Weaver, Zoe
: TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cancer
: TITLE OF INVENTION: Gene Sets
: FILE REFERENCE: 689290-77
: CURRENT APPLICATION NUMBER: US/09/954,531
: CURRENT FILING DATE: 2002-05-02
: PRIOR APPLICATION NUMBER: US/60/233,133
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US/60/234,009
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,034
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,509
: PRIOR FILING DATE: 2000-09-22
: PRIOR APPLICATION NUMBER: US/60/234,567
: PRIOR FILING DATE: 2000-09-22
: NUMBER OF SEQ ID NOS: 1392
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 140
: LENGTH: 2439
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-954-531-140

```

```

Query Match 38.3%; Score 142.6; DB 10; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4, 6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 98 GTTCAGACTGTGGCCCCGAGATGATCTTCGGGAACTGCAGAGAAACCAACGCGCGCCTGCAGG 157
Db 105 GCTCAGACTGTGGCCCCGAGATGATCTTCGGGAACTGCAGAGAAACCAACGCGCGCCTGCAGG 164
OY 158 ACGTCCGGAGCTGCTGTGGCGCAGCAGAGTCAGGAGATCAACGTTCTGAAAAACACGGTGA 217
Db 165 ACGTCCGGAGCTGCTGTGGCGCAGCAGATCAGGAGAGATCAAGTTCTGAAAAACACGGTGA 224
OY 218 TGGAGTGTGACGCGGTGCGGGCCGACGCCG 246
Db 225 TGGAGTGTGACGCGGTGCGGGAGATGCAGCAG 253

RESULT 4
US-09-954-531-359
; Sequence 359, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 359
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-359

Query Match 38.3%; Score 142.6; DB 10; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4, 6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 98 GTTCAGACTGTGGCCCCGAGATGATCTTCGGGAACTGCAGAGAAACCAACGCGCGCCTGCAGG 157
Db 105 GCTCAGACTGTGGCCCCGAGATGATCTTCGGGAACTGCAGAGAAACCAACGCGCGCCTGCAGG 164
OY 158 ACGTCCGGAGCTGCTGTGGCGCAGCAGATCAGGAGATCAACGTTCTGAAAAACACGGTGA 217
Db 165 ACGTCCGGAGCTGCTGTGGCGCAGCAGATCAGGAGAGATCAAGTTCTGAAAAACACGGTGA 224
OY 218 TGGAGTGTGACGCGGTGCGGGCCGACGCCG 246
Db 225 TGGAGTGTGACGCGGTGCGGGAGATGCAGCAG 253

RESULT 5
US-09-918-624B-49
; Sequence 49, Application US/09918624B
; Publication No. US20030113720A1
; GENERAL INFORMATION:
; APPLICANT: Scheyde, Xiao Min
; APPLICANT: Sothase, Thierly
; TITLE OF INVENTION: CDNAS EXPRESSED IN ADIPOCYTE DIFFERENTIATION
; FILE REFERENCE: PA-0033 US

```

```

CURRENT APPLICATION NUMBER: US/09/918,624B
CURRENT FILING DATE: 2002-12-03
PRIOR APPLICATION NUMBER: 60/222,470
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PERL Program
SEQ ID NO 49
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: GenBank ID No. US20030113720A1 9602449
US-09-918-624B-49

Query Match      38.3%; Score 142.6; DB 11; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4.6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGGCTGCAGG 157
Db 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGGCTGCAGG 164

Oy 158 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCAGCTTCTGAAAAACACGGTGA 217
Db 165 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCAGCTTCTGAAAAACACGGTGA 224

Oy 218 TGGAGTGTGACGCGTGGCGGCGCGGCGGCGG 246
Db 225 TGGAGTGTGACGCGTGGCGGAGTGCAGCAG 253

RESULT 6
US-10-301-822-40
Sequence 40, Application US/10301822
Publication No. US20030148410A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
APPLICANT: Burgart, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MPM01-029P2RMM
CURRENT APPLICATION NUMBER: US/10/301,822
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 40
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (26) ... (2299)
US-10-301-822-40

Query Match      38.3%; Score 142.6; DB 12; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4.6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGGCTGCAGG 157
```

```

Db 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGGCTGCAGG 164

Oy 158 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCAGCTTCTGAAAAACACGGTGA 217
Db 165 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCAGCTTCTGAAAAACACGGTGA 224

Oy 218 TGGAGTGTGACGCGTGGCGGCGCGGCGGCGGCGG 246
Db 225 TGGAGTGTGACGCGTGGCGGAGTGCAGCAG 253

RESULT 7
US-10-177-293-71
Sequence 71, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Ganavarrpu, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Meyer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
PRIOR FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 71
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
US-10-177-293-71

Query Match      38.3%; Score 142.6; DB 14; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4.6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGGCTGCAGG 157
Db 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGGCTGCAGG 164

Oy 158 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCAGCTTCTGAAAAACACGGTGA 217
Db 165 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCAGCTTCTGAAAAACACGGTGA 224

Oy 218 TGGAGTGTGACGCGTGGCGGCGCGGCGGCGGCGG 246
```

Db 225 TGGAGTGTACCGCTGCGGAGATGACAGCAG 253

RESULT 8
US-09-111-911-5/c

/ Sequence 5, Application US/09111911
/ Publication No. US20030096768A1
/ GENERAL INFORMATION:
/ APPLICANT: Wold, William S.M.
/ TITLE OF INVENTION: Inhibiting Apoptosis With Adenovirus R1D Protein
/ FILE REFERENCE: 16153-5587
/ CURRENT FILING DATE: 1998-07-08
/ NUMBER OF SEQ ID NOS: 5
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 34427
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Plasmid
/ OTHER INFORMATION: Combining E.coli and Adenovirus Sequences
US-09-111-911-5

Query Match 15.6%; Score 58; DB 11; Length 34427;
Best Local Similarity 92.4%; Pred. No. 5e-07;
Matches 61; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 CCATGGGAAGGTACATGATTTTAAAGCTTGCTGCGCCCTTGCGGAGTCTGCGAGCGCTGCCA 67
Db 2656 CCAAGATGAGTACATGATTTTAAAGCTTGCTGCGCCCTTGCGGAGTCTGCGAGCGCTGCCA 2597

QY 68 AAAAG 73
Db 2596 AAAAG 2591

RESULT 9

US-09-782-378A-3
/ Sequence 3, Application US/09782378A
/ Patent No. US20020102731A1
/ GENERAL INFORMATION:
/ APPLICANT: Hearing, Patrick
/ APPLICANT: Bahou, Wadie
/ APPLICANT: Sandalon, Ziv
/ APPLICANT: Gnatenko, Dmitrii
/ TITLE OF INVENTION: Adenoviral Vectors
/ FILE REFERENCE: STONY-04970
/ CURRENT APPLICATION NUMBER: US/09/782,378A
/ CURRENT FILING DATE: 2001-02-12
/ PRIOR APPLICATION NUMBER: 60/237,747
/ PRIOR FILING DATE: 2000-10-02
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: Patent In version 3.0
/ SEQ ID NO 3
/ LENGTH: 35937
/ TYPE: DNA
/ ORGANISM: Human adeno-associated virus 2
US-09-782-378A-3

Query Match 15.6%; Score 58; DB 10; Length 35937;
Best Local Similarity 92.4%; Pred. No. 5.1e-07;
Matches 61; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 CCATGGGAAGGTACATGATTTTAAAGCTTGCTGCGCCCTTGCGGAGTCTGCGAGCGCTGCCA 67
Db 28807 CCAAGATGAGTACATGATTTTAAAGCTTGCTGCGCCCTTGCGGAGTCTGCGAGCGCTGCCA 28866

QY 68 AAAAG 73
Db 28867 AAAAG 28872

RESULT 10
US-10-156-761-4994

/ Sequence 4994, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 4994
/ LENGTH: 1734
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1734)
US-10-156-761-4994

Query Match 15.5%; Score 57.8; DB 14; Length 1734;
Best Local Similarity 89.9%; Pred. No. 3.1e-07;
Matches 62; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 237 GCCGAGCCGCGAAGCCGCAAGCCGCGAGCCGCGAGCCGCGAGCCGCGAAGCCGCGA 296
Db 105 GCCGAGCCGCGAAGCCGCGAAGCCGCGAGCCGCGAGCCGCGAGCCGCGAAGCCGCGA 164

QY 297 ACCGGAACC 305
Db 165 GCCGCAACC 173

RESULT 11

US-10-156-761-1
/ Sequence 1, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 1
/ LENGTH: 9025608
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (4187715)
/ OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Query Match 15.5%; Score 57.8; DB 14; Length 9025608;

Best Local Similarity 89.9%; Pred. No. 1.8e-06;

Matches 62; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 237 GCCGAGCCGAGCCGGAACCGCAGCCGCGAGCCGCGAGCCGGAACCGCAGCCGGA 296

Db 6065691 GCCGAGCCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGCA 6065750

QY 297 ACCGGAAC 305

Db 6065751 GCCGGAAC 6065759

RESULT 12

US-09-918-995-4559

Sequence 4559, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4559

LENGTH: 428

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)...(428)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-4559

Query Match 14.5%; Score 53.8; DB 11; Length 428;

Best Local Similarity 55.7%; Pred. No. 3.5e-06;

Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 116 AGATGCTTCGGGAATGAGAAACCGCGCGCTGCGAGAGCTGCGGAGCTGCTGC 175

Db 68 AGATGCTGAGGCGCATCTGCGGTCAATCCAGCGCGAGAGAGAGAGAGAGAGAGAGAG 127

QY 176 GCGAGAGGTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 235

Db 128 AGCAG 187

QY 236 GCGCGAG 295

Db 188 GCGAG 247

QY 296 AACCG 300

Db 248 AGCAG 252

RESULT 13

US-09-918-995-4560

Sequence 4560, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4560

LENGTH: 433

TYPE: DNA

ORGANISM: Homo sapiens

US-09-918-995-4560

Query Match 14.5%; Score 53.8; DB 11; Length 433;

Best Local Similarity 55.7%; Pred. No. 3.5e-06;

Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 116 AGATGCTTCGGGAATGAGAAACCGCGCGCTGCGAGAGAGCTGCGGAGCTGCTGC 175

Db 68 AGATGCTGAGGCGCATCTGCGGTCAATCCAGCGCGAGAGAGAGAGAGAGAGAGAGAG 127

QY 176 GCGAGAGGTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 235

Db 128 AGCAG 187

QY 236 GCGCGAG 295

Db 188 GCGAG 247

QY 296 AACCG 300

Db 248 AGCAG 252

RESULT 14

US-09-735-705-324

Sequence 324, Application US/09735705

Patent No. US20020052329A1

GENERAL INFORMATION:

APPLICANT: Wang, Tongtong

APPLICANT: Fan, Liqun

APPLICANT: Kailos, Michael D.

APPLICANT: Bangur, Chaitanya S.

APPLICANT: Hosken, Nancy

APPLICANT: Fanger, Gary R.

APPLICANT: Li, Samuel X.

APPLICANT: Wang, Aijun

APPLICANT: Skelky, Yasir A.W.

APPLICANT: Henderson, Robert A.

APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.455C14

CURRENT APPLICATION NUMBER: US/09/735,705

CURRENT FILING DATE: 2000-12-12

NUMBER OF SEQ ID NOS: 419

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 324

LENGTH: 521

TYPE: DNA

ORGANISM: Homo sapien

US-09-735-705-324

Query Match 14.2%; Score 53; DB 9; Length 521;

Best Local Similarity 58.6%; Pred. No. 6.3e-06;

Matches 92; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 149 CGCTGAGAGAGTGGAGTGGCTGCGGAGAGAGTCAAGGAGATGACGTTCTGAAA 208

Db 52 CGCTGAGAGAGTGGAGTGGCTGCGGAGAGAGTCAAGGAGATGACGTTCTGAAA 111

QY 209 ACAGGAGAGAGTGGAGTGGCTGCGGAGAGAGTCAAGGAGATGACGTTCTGAAA 268

Db 112 AGAAGAGAGAGTGGAGTGGCTGCGGAGAGAGTCAAGGAGATGACGTTCTGAAA 171

QY 269 GCGAGAGAGAGTGGAGTGGCTGCGGAGAGAGTCAAGGAGATGACGTTCTGAAA 305

Db 172 TCCAGCCCCAACCCGAGCTCTCAGCTCTCAGCCGCAACC 208

RESULT 15

US-09-850-716A-324
; Sequence 324, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
; APPLICANT: Kalos, Michael D.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Retter, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C15
; CURRENT APPLICATION NUMBER: US/09/850,716A
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 324
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-850-716A-324

Query Match 14.2%; Score 53; DB 10; Length 521;
Best Local Similarity 58.6%; Pred. No. 6.3e-06;
Matches 92; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 149 CGCTGACGAGCTGCGGAGCTGCTCGGACGACAGTCAAGGAGATCACGTTCTGAAAA 208
DB 52 CGCTGACAGATGTGTGACAGTACAGAAATCGTCAAGCCATCTGTGGGTCAATCCACGCGGC 111
QY 209 ACACGGATGTGAGTGTGACGCGTGGGCGGAGCCGCAAGCCGAAACCGACCGCGACG 268
DB 112 AGAAGCAGACGACCTGTCTTCCAGCAGACCCCTCTCGCAGCGGACGCGCAGCGACG 171
QY 269 CGCAGCCGACCGCAACCGCAGCCGCAAGCCGAAACCGGAAC 305
DB 172 TCCAGCCCCAACCCGAGCTCTCAGCTCTCAGCCGCAACC 208

Search completed: August 26, 2003, 16:55:26
Job time : 822 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:10:25 ; Search time 1432 Seconds

(without alignments)
6113.733 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372
Sequence: 1 aagcttaccatggaagta.....aatccatgcatctcgag 372

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_eebba: *
2: em_eebba: *
3: em_eebba: *
4: em_eebba: *
5: em_eebba: *
6: em_eebba: *
7: em_eebba: *
8: em_eebba: *
9: em_eebba: *
10: em_eebba: *
11: em_eebba: *
12: em_eebba: *
13: em_eebba: *
14: em_eebba: *
15: em_eebba: *
16: em_eebba: *
17: em_eebba: *
18: em_eebba: *
19: em_eebba: *
20: em_eebba: *
21: em_eebba: *
22: em_eebba: *
23: em_eebba: *
24: em_eebba: *
25: em_eebba: *
26: em_eebba: *
27: em_eebba: *
28: em_eebba: *
29: em_eebba: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match length	ID	Description
1	137.8	37.0	328 9	AT751993 cn13a10.x
2	137.8	37.0	365 12	BG900764 HOA42-1-H
3	137.8	37.0	366 12	BG897326 HOA12-1-E
4	137.8	37.0	370 12	BG896418 HOA30-1-F

5	137.8	37.0	370 12	BG897214 HOA33-1-F
6	137.8	37.0	371 12	BG898406 HOA8-1-F1
7	137.8	37.0	372 12	BG899727 HOA3-1-A
8	137.8	37.0	373 12	BG896387 HOA30-1-E
9	137.8	37.0	374 12	BG896390 HOA30-1-E
10	137.8	37.0	375 12	BG900556 HOA42-1-C
11	137.8	37.0	376 12	BG900556 HOA42-1-C
12	137.8	37.0	377 12	BG898027 HOA20-1-H
13	137.8	37.0	378 12	BG896332 HOA29-1-D
14	137.8	37.0	379 12	BG900343 HOA37-1-F
15	137.8	37.0	380 12	BG898382 HOA8-1-F4
16	137.8	37.0	381 12	BG896503 HOA30-1-C
17	137.8	37.0	382 12	BG899936 HOA40-1-E
18	137.8	37.0	383 12	BG900459 HOA41-1-C
19	137.8	37.0	384 12	BG897369 HOA12-1-B
20	137.8	37.0	385 12	BG897511 HOA14-1-A
21	137.8	37.0	386 12	BG898968 HOA21-1-C
22	137.8	37.0	387 12	BG900465 HOA41-1-C
23	137.8	37.0	388 12	BG898309 HOA39-1-A
24	137.8	37.0	389 12	BG898604 HOA19-1-G
25	137.8	37.0	390 12	BG898604 HOA19-1-G
26	137.8	37.0	391 12	BG896715 HOA49-1-D
27	137.8	37.0	392 12	BG897500 HOA14-1-G
28	137.8	37.0	393 12	BG898326 HOA55-1-C
29	137.8	37.0	394 12	BG898329 HOA55-1-D
30	137.8	37.0	395 12	AT751616 cn11a01.Y
31	137.8	37.0	396 12	BG898451 HOA9-1-D1
32	137.8	37.0	397 12	AM067829 cn19f09.X
33	137.8	37.0	398 12	BG898353 HOA55-1-D
34	137.8	37.0	399 12	BG900134 HOA51-1-D
35	137.8	37.0	400 12	BG898861 HOA40-1-C
36	137.8	37.0	401 12	BG898320 HOA55-1-C
37	137.8	37.0	402 12	BG897459 HOA13-1-H
38	137.8	37.0	403 12	BG898924 HOA40-1-F
39	137.8	37.0	404 12	BG896555 HOA47-1-D
40	137.8	37.0	405 12	BG901017 HOA52-1-G
41	137.8	37.0	406 12	BG901170 HOA54-1-H
42	137.8	37.0	407 12	AU298537 AU298537
43	137.8	37.0	408 12	BG899058 HOA21-1-E
44	137.8	37.0	409 12	BG899411 HOA45-1-A
45	137.8	37.0	410 12	BG899174 HOA22-1-D

ALIGNMENTS

RESULT 1
LOCUS AT751993 328 bp mRNA linear EST 20-JUN-2002
DEFINITION cn13a10.x1 Normal Human Trabecular Bone Cells Homo sapiens CDNA
clone NHTBC_cn13a10 random, mRNA sequence.

ACCESSION AT751993
VERSION AT751993.1 GI:5130257

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-402-4877
Fax: 301-496-7157
Email: libin@nih.gov
DNA Sequencing and analyses by National Institutes of Health
Intramural Sequencing Center (NISC).

Plate: 13 row: a column: 10
 Seq primer: -21M13 forward primer (ABI).
 Location/Qualifiers

1. 328

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="NHTBC.cn13a10"

/sex="Female"

/tissue_type="Bone"

/cell_type="Trabecular Bone Cells"

/clone_lib="Normal Human Trabecular Bone Cells"

/note="Organ: Hip; Vector: pBluescript; Site_1: EcoRI; library constructed by Dr. Marian Young and Dr. Pamela Gehron Robey (NIDCR)"

BASE COUNT 51 a 119 c 108 g 50 t

ORIGIN

Query Match 37.0%; Score 137.8; DB 9; Length 328;

Best Local Similarity 95.3%; Pred. No. 4.1e-22;

Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 157
 DB 108 GCTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 167
 QY 158 AGTCGGGAGCTGGGCGGCGAGAGTCAAGGAGATCAGTTCCGAAAAACACGGTGA 217
 DB 168 AGTCGGGAGCTGGGCGGCGAGAGTCAAGGAGATCAGTTCCGAAAAACACGGTGA 227
 QY 218 TGGAGTGTGACCGCTGCGGCGCGAGCCG 246
 DB 228 TGGAGTGTGACCGCTGCGGCGAGATGCAGCAG 256

RESULT 2
 LOCUS BG900764 365 bp mRNA linear EST 06-NOV-2001
 DEFINITION HOA42-1-H5 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
 mRNA sequence.

ACCESSION BG900764
 VERSION BG900764.1 GI:14311013
 KEYWORDS EST.

SOURCE
 ORGANISM Homo sapiens (human)

COMMENT
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 365)

Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
 Sathie,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
 Lark,M.W.

Identification and initial characterization of 5000 expressed
 sequenced tags (ESTs) each from adult human normal and
 osteoarthritic cartilage cDNA libraries

Osteoarthr. Cartil. 9 (7), 641-653 (2001)

21482651

11597177

CONTACT: Sanjay Kumar

UM2109

GlaxoSmithKline

709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA

Tel: 610-270-7245

Fax: 610-270-5598

Email: sanjay.kumar-1@gsk.com

Seq primer: T7.

Location/Qualifiers

1. 365

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="cartilage"

/lab_host="E.coli DH10 B"

/clone_lib="HOA (Human Osteoarthritic Cartilage)"
 /note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI;
 Directional"

BASE COUNT 65 a 123 c 121 g 56 t

ORIGIN

Query Match 37.0%; Score 137.8; DB 12; Length 365;

Best Local Similarity 95.3%; Pred. No. 4.2e-22;

Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 157
 DB 160 GCTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 239
 QY 158 AGTCGGGAGCTGGGCGGCGAGAGTCAAGGAGATCAGTTCCGAAAAACACGGTGA 217
 DB 240 AGTCGGGAGCTGGGCGGCGAGAGTCAAGGAGATCAGTTCCGAAAAACACGGTGA 299
 QY 218 TGGAGTGTGACCGCTGCGGCGCGAGCCG 246
 DB 300 TGGAGTGTGACCGCTGCGGCGAGATGCAGCAG 328

RESULT 3
 LOCUS BG897326 366 bp mRNA linear EST 06-NOV-2001
 DEFINITION HOA12-1-E3 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
 mRNA sequence.

ACCESSION BG897326
 VERSION BG897326.1 GI:14307567
 KEYWORDS EST.

SOURCE
 ORGANISM Homo sapiens (human)

COMMENT
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 366)

Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
 Sathie,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
 Lark,M.W.

Identification and initial characterization of 5000 expressed
 sequenced tags (ESTs) each from adult human normal and
 osteoarthritic cartilage cDNA libraries

Osteoarthr. Cartil. 9 (7), 641-653 (2001)

21482651

11597177

CONTACT: Sanjay Kumar

UM2109

GlaxoSmithKline

709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA

Tel: 610-270-7245

Fax: 610-270-5598

Email: sanjay.kumar-1@gsk.com

Seq primer: T7.

Location/Qualifiers

1. 366

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="cartilage"

/lab_host="E.coli DH10 B"

/clone_lib="HOA (Human Osteoarthritic Cartilage)"

/note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI;
 Directional"

BASE COUNT 56 a 134 c 117 g 59 t

ORIGIN

Query Match 37.0%; Score 137.8; DB 12; Length 366;

Best Local Similarity 95.3%; Pred. No. 4.2e-22;

Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 157
 DB 148 GCTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTGCAGG 207

Oy	158	ACGTCGGAGCATCGGTGGCGGAGCAGCGATCAAGTCTCTGAAAAAACCAGTGGA	217
Dd	208	ACGTCGGGAGACTGCTGCGCCAGCAGGATCAAGGAGATCATCTTCTGAAAAAACAGTGGA	267
Oy	218	TGGAAGTGTGACCGCGTGCAGGCGCCGACGCCG	246
Dd	268	TGGAAGTGTGACCGCGTGCAGGATGACGACG	296
 RESULT 4			
LOCUS	BG896418	370 bp	mRNA linear EST 06-NOV-2001
DEFINITION	HOA30-1-F11 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,		
ACCESSION	BG896418		
VERSION	BG896418.1	GI:14306659	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	I (bases 1 to 370) Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J., Sache,G., Muir,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.		
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries		
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)		
MEDLINE	21482651		
PUBMED	11597177		
COMMENT	Contact: Sanjay Kumar UM2109		
FEATURES	Glaxosmithkline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel.: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@gsk.com Seq primer: T7. Location/Qualifiers 1..370 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /tissue_type="cartilage" /lab_host="E.coli DH10 B" /clone_id="HOA (Human Osteoarthritic Cartilage)" /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI; Directional"		
BASE COUNT	68 a 128 c 118 g 56 t		
ORIGIN			
Query Match	37.0%; Score 137.8; DB 12; Length 370;		
Best Local Similarity	95.3%; Pred. No. 4,2e-22;		
Matches 142; Conservative	0; Mismatches 7; Indels 0; Gaps 0;		
Oy	98	GTTGAGACTGTGGGCCCCGAGATGCTTGGGGAACTGACAGAAAACAAGCGCGCTGCAGG	157
Dd	189	GCTCAGACTGTGGCCCCGAGATGCTTGGGAACTGACAGAAAACAAGCGCGCTGCAGG	248
Oy	158	ACGTGCGGAGACTGTGCTGCGGAGCAGGTCAAGGAGATCATCGTTCTGAAAAAACAGTGTA	217
Dd	249	ACGTGCGGAGAGTGTGCTGCGGAGCAGGTCAGGAGATCATCGTTCTGAAAAAACAGTGTA	308
Oy	218	TGGAAGTGTGACCGCGTGCAGGCGCCGACGCCG	246
Dd	309	TGGAAGTGTGACCGCGTGCAGGATGACGACG	337
 RESULT 5			
LOCUS	BG897214	370 bp	mRNA linear EST 06-NOV-2001
	BG897214		

DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT	FEATURES	
HOA33-1-F1 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.	BC897214	BC897214.1	GI:14307455	Homo sapiens (human)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 370)	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathie,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries	Osteoarthr. Cartil. 9 (7), 641-653 (2001)	21482651	11597177	Contact: Sanjay Kumar UW2109 Glaxosmithkline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay_kumar-1@sk.com Seq primer: T7.	Location/Qualifiers 1..370 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /tissue_type="cartilage" /lab_host="E.coli DH10 B" /clone_id="HOA (Human Osteoarthritic Cartilage)" /note=Vector: pSPORT I; Site_1: SalI; Site_2: NotI; Directional"	
Query Match	37.0%;	Score 137.8;	DB 12;	Length 370;	Best Local Similarity	95.3%;	Pred. No. 4,2e-22;	Matches 142;	Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0;
Db	98	GTTCAGACCTGGGCGCCGACAGTGGCTTGGGGAATGACAGAAACCAACGCGCGCTGCAGG	157										
Db	189	GCTAGACCTGGGCGCCGACAGTGGCTTGGGGAATGACAGAAACCAACGCGCGCTGCAGG	248										
Qy	158	ACGTCGGGAGCTGCTGCGCGACGAGTCAAGGAGATCAAGTCTCTGAAAAACACGGTGA	217										
Db	249	ACGTCGGGAGCTGCTGCGCGACGAGTCAAGGAGATCAAGTCTCTGAAAAACACGGTGA	308										
Qy	218	TGAGTGTGAACGGCTGCGGCGCGAGCGG	246										
Db	309	TGAGTGTGAACGGCTGCGGCGAGTGCAGCAG	337										
RESULT 6	BC898406	BC898406	371 bp	mRNA	linear	EST 06-NOV-2001							
LOCUS	HOA8-1-F10 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.	BC898406	BC898406.1	GI:14308655	Homo sapiens (human)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 371)	Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathie,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.						

TITLE	IDENTIFICATION AND INITIAL CHARACTERIZATION OF 5000 EXPRESSED SEQUENCED TAGS (ESTs) EACH FROM ADULT HUMAN NORMAL AND OSTEOARTHRITIC CARTILAGE CDNA LIBRARIES
JOURNAL	OSTEOARTHRITIC CARTILAGE CDNA LIBRARIES
MEDLINE	OSTEOARTHR. CARTIL. 9 (7), 641-653 (2001)
PUBMED	21482651
COMMENT	11597177 Contact: Sanjay Kumar UM2109 GlaxoSmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@gsk.com Seq primer: T7.
FEATURES	Location/Qualifiers
SOURCE	1..371 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /issue_type="cartilage" /lab_host="E.coli DH10 B" /clone_lib="HOA (Human Osteoarthritic Cartilage)" /note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI; Directional"
BASE COUNT	65 a 126 c 118 g 62 t
ORIGIN	
Query Match	37.0% Score 137.8; DB 12; Length 371;
Best Local Similarity	95.3% Pred. No. 4.2e-22;
Matches 142;	Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY	98 GTTCAGACCTTGCGGCCCGCAGATGCTTGCGGAACTGCAGAAACCAACGCGCGCTGCAGG 157
Db	157 GCTCAGACCTTGCGGCCCGCAGATGCTTGCGGAACTGCAGAAACCAACGCGCGCTGCAGG 216
QY	158 ACGTGCAGGAGCTGCTGCGGAGCAGACGTCAGGGAGATCAGCTTCTGAAAAACCGGTGA 217
Db	217 ACGTGCAGGAGCTGCTGCGGAGCAGACGTCAGGGAGATCAGCTTCTGAAAAACCGGTGA 276
QY	218 TGAAGTGTGACGCGTGCAGGCGCGCAGCCG 246
Db	277 TGAAGTGTGACGCGTGCAGGCGCGCAGCCG 305
RESULT 7	372 bp mRNA linear EST 06-NOV-2001
LOCUS	HOA32-1-A8 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.
ACCESSION	BC899727
VERSION	BC899727.1 GI:14309992
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 372) Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathé,G., Muti,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)
MEDLINE	21482651
PUBMED	11597177
COMMENT	Contact: Sanjay Kumar UM2109 GlaxoSmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@gsk.com

FEATURES		Seq primer: T7.		Location/Qualifiers	
SOURCE		1..372			
BASE COUNT		68 a	130 c	117 g	57 t
ORIGIN					
Query Match		37.0%;	Score 137.8;	DB 12;	Length 372;
Best Local Similarity		95.3%;	Pred. No. 4.2e-22;		
Matches 142;		Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
Qy	98 GTTCAGACCTGGGCGCCGACAGATGCTTGGGAACTGCAGAAACCAACGCGCGCTGCAGG	157			
Db	189 GCTCAGACCTGGGCGCCGACAGATGCTTGGGAACTGCAGAAACCAACGCGCGCTGCAGG	248			
Qy	158 ACGTCCGGACCTGGCTTCGGGACAGAGTCAGGAGATCAAGTTCTGAAAAACACGGTGA	217			
Db	249 ACGTCCGGACCTGGCTTCGGGACAGAGTCAGGAGATCAAGTTCTGAAAAACACGGTGA	308			
Qy	218 TGGAGTGTGACGCGTCGCGGCGCCGACGCG	246			
Db	309 TGGAGTGTGACGCGTCGCGGCGCCGACGCG	337			
<p>RESULT 8 BG896387 LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM</p> <p> BG896387 HOA30-1-87 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA, sequence. BG896387 BG896387.1 GI:14306628 EST. Homo sapiens (human) Homo sapiens Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Cetartihmi; Homnidae; Homo. 1 (bases 1 to 372) Kumar,S., Connor,U.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathie,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W. Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries Osteoarthr. Cartil. 9 (7), 641-653 (2001) 21482651 11597177 Contact: Sanjay Kumar UW2109 GlaxoSmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@gsk.com Seq primer: T7.</p>					
<p>FEATURES</p> <p>SOURCE</p> <p>1..372</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/cruise_type="cartilage"</p> <p>/lab_host="E.coli DH10 B"</p> <p>/clone_lid="HOA (Human Osteoarthritic Cartilage)"</p> <p>/note="Vector: pSPORT I, Site_1: SalI, Site_2: NotI; Directional"</p>					
<p>BASE COUNT</p> <p>68 a 130 c 117 g 57 t</p> <p>ORIGIN</p>					

Query Match 37.0%; Score 137.8; DB 12; Length 373;
 Best Local Similarity 95.3%; Pred. No. 4.2e-22;
 Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 157
 DB 189 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 248
 QY 158 ACGTGGCGGAGCTGCTGGCGGAGAGATGATCACTTCCTGAAAAACCGGTGA 217
 DB 249 ACGTGGCGGAGCTGCTGGCGGAGAGATGATCACTTCCTGAAAAACCGGTGA 308
 QY 218 TGGAGTGTGACGCGTGCAGCGCGCGAGCCG 246
 DB 309 TGGAGTGTGACGCGTGCAGCGCGAGTGCAGCAG 337

RESULT 9 BG896390 373 bp mRNA linear EST 06-NOV-2001
 LOCUS HOA30-1-B10 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
 DEFINITION mRNA sequence.
 ACCESSION BG896390
 VERSION BG896390.1 GI:14306631
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 373)
 Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
 Sathes,G., Mu,J.P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
 Lack,M.W.
 Identification and initial characterization of 5000 expressed
 sequenced tags (ESTs) each from adult human normal and
 osteoarthritic cartilage cDNA libraries
 Osteoarthr. Cartil. 9 (7), 641-653 (2001)
 21482651
 11597177
 COMMENT Contact: Sanjay Kumar
 UW2109
 Glaxosmithkline
 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
 Tel: 610-270-7245
 Fax: 610-270-5598
 Email: sanjay_kumar-1@gsk.com
 Seq primer: 17.
 Location/Qualifiers
 1..373
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="cartilage"
 /lab_host="E.coli DH10 B"
 /clone_lib="HOA (Human Osteoarthritic Cartilage)"
 /note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI;
 Directional"
 BASE COUNT 69 a 129 c 118 g 57 t
 ORIGIN

Query Match 37.0%; Score 137.8; DB 12; Length 373;
 Best Local Similarity 95.3%; Pred. No. 4.2e-22;
 Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 157
 DB 190 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 249
 QY 158 ACGTGGCGGAGCTGCTGGCGGAGAGATGATCACTTCCTGAAAAACCGGTGA 217
 DB 250 ACGTGGCGGAGCTGCTGGCGGAGAGATGATCACTTCCTGAAAAACCGGTGA 309

QY 218 TGGAGTGTGACGCGTGCAGCGCCGAGCCG 246
 DB 310 TGGAGTGTGACGCGTGCAGGATGCAGCAG 338

RESULT 10 AM068338 374 bp mRNA linear EST 20-JUN-2002
 LOCUS cn20g10.y1 Normal Human Trabecular Bone Cells Homo sapiens cDNA
 DEFINITION clone NHBC_cn20g10 random, mRNA sequence.
 ACCESSION AM068338
 VERSION AM068338.1 GI:6023336
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 374)
 Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,
 Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M., Robey
 P.G., Hotchkiss,R.N. and Francomano,C.A.
 SCAP: The Skeletal Genome Anatomy Project
 Unpublished
 TITLE JOURNAL
 COMMENT Contact: Libin Jia
 Medical Genetics Branch
 National Human Genome Research Institute
 10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
 Tel: 301-402-4877
 Fax: 301-496-7157
 Email: libin@helix.nih.gov
 DNA Sequencing and analyses by National Institutes of Health
 Intramural Sequencing Center (NISC).
 Plate: 20 row: 9 column: 10
 Seq primer: M13Rp1 reverse primer (ABI).
 Location/Qualifiers
 1..374
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NHBC_cn20g10"
 /sex="Female"
 /tissue_type="Bone"
 /cell_type="Trabecular Bone Cells"
 /lab_host="SURE"
 /clone_lib="Normal Human Trabecular Bone Cells"
 /note="Organ: Hip; Vector: pBlueScript; Site_1: EcoRI;
 Library constructed by Dr. Marian Young and Dr. Pamela
 Gehron Robey (NIDCR)"
 BASE COUNT 55 a 138 c 124 g 57 t
 ORIGIN

Query Match 37.0%; Score 137.8; DB 9; Length 374;
 Best Local Similarity 95.3%; Pred. No. 4.2e-22;
 Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 157
 DB 103 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 162
 QY 158 ACGTGGCGGAGCTGCTGGCGGAGAGATGATCACTTCCTGAAAAACCGGTGA 217
 DB 163 ACGTGGCGGAGCTGCTGGCGGAGAGATGATCACTTCCTGAAAAACCGGTGA 222
 QY 218 TGGAGTGTGACGCGTGCAGCGCGCGAGCCG 246
 DB 223 TGGAGTGTGACGCGTGCAGCGCGAGTGCAGCAG 251

RESULT 11 BG900556 375 bp mRNA linear EST 06-NOV-2001
 LOCUS HOA42-1-C6 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
 DEFINITION mRNA sequence.

FEATURES	Source	Location/Qualifiers
BASE COUNT	67 a 130 c 120 g 58 t	
ORIGIN		
Query Match	37.0%; Score 137.8; DB 12; Length 375;	
Best Local Similarity	95.3%; Pred. No. 4.2e-22;	
Matches 142; Conservative	0; Mismatches 7; Indels 0; Gaps 0;	
QY	98 GTTACAGACTCTGGGCGCCGACATGCTTGGGGAATCTGCAGAAACCAACCGCGCTGCAGG	157
Db	177 GCTCAGACCTGGGCGCCGACATGCTTGGGGAATCTGCAGAAACCAACCGCGCTGCAGG	236
QY	158 ACGTCCGGGACCTGCTGCGGCGAGCAGTGCAGGAGATCAGCTTCTGAAAAACACCGTGA	217
Db	237 ACGTCCGGGACCTGCTGCGGCGAGCAGTGCAGGAGATCAGCTTCTGAAAAACACCGTGA	296
QY	218 TGGAGTGTGACGCGTGCAGGCGGACCGG	246
Db	297 TGGAGTGTGACGCGTGCAGGCGGATGCAGCG	325
RESULT 12		
LOCUS	BC898027	
DEFINITION	HOA20-1-13 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,	
VERSION	BC898027	
KEYWORDS	EST.	
ORGANISM	Homo sapiens (human)	
REFERENCE	1 (bases 1 to 375)	
AUTHORS	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Satche,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.	
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries	
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)	
MEDLINE	21482651	
PUBMED	11597177	
COMMENT	Contact: Sanjay Kumar UM2109 GlaxoSmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@sk.com Seq primer: T7	
FEATURES		
Source	1..375	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/feature_type="cartilage"	
	/lab_host="E.coli DH10 B"	
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"	
	/note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI; Directional"	

JOURNAL	Osteoarthritic cartilage cDNA libraries
MEDLINE	21482651
PUBMED	11597177
COMMENT	Contact: Sanjay Kumar UM2109 GlaxosmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel.: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@gsk.com Seq primer: T7.
FEATURES	Location/Qualifiers
SOURCE	1..379 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /lab_type="cartilage" /lig_host="E.coli DH10 B" /clone_id="HOA (Human Osteoarthritic Cartilage)" /note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI; Directional"
BASE COUNT	67 a. 134 c 119 g 59 t
ORIGIN	C
Query Match	37.0%; Score 137.8; DB 12; Length 379;
Best Local Similarity	95.3%; Pred. No. 4.2e-22;
Matches 142; Conservative	0; Mismatches 7; Indels 0; Gaps 0;
Dy	98 GTTCAGACCTGGGCGCCGACATGCTTGGGGAACTGCAGAAACCAAGCGCGGTGCAGG 157
Dd	181 GCTAGACCTGGGCGCCGACATGCTTGGGAATGTCAGAAACCAGCGCGCTGCAGG 240
Dy	158 ACGTGCCGGA CTGGCTCGCGCAGCAGGTGCAGGAGATCACGTTCTGAAAAACACGGTGA 217
Dd	241 ACGTGCGGAGCTGCTCGCGCAGCAGGTGCAGGAGATCACGTTCTGAAAAACACGGTGA 300
Dy	218 TGAGTGTGACGCGTGGCGCGCGCGG 246
Dd	301 TGGAGTGTGACGCGTGGCGCGCATGCAGCG 329
RESULT 13	
Bg896332	380 bp mRNA linear EST 06-NOV-2001
LOCUS	HOA29-1-112 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, definition
ACCESSION	BG896332
VERSION	BG896332.1 GI:14306573
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 380) Kumar,S., Connor,J.R., Dodde,R.A., Halsey,W., Van Horn,M., Mao,J., Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.
AUTHORS	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries Osteoarthr. Cartil. 9 (7), 641-653 (2001)
TITLE	
JOURNAL	
MEDLINE	
PUBMED	11597177
COMMENT	Contact: Sanjay Kumar UM2109 GlaxosmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel.: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@gsk.com Seq primer: T7. Location/Qualifiers
FEATURES	

source
1. .380
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/clone_lib="HOA (Human Osteoarthritic Cartilage)"
/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 67 a 134 c 120 g 59 t
ORIGIN

Query Match 37.0%; Score 137.8; DB 12; Length 380;
Best Local Similarity 95.3%; Pred. No. 4.2e-22;
Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 98 GTTACAGCTGGGCGCCGAGATGCTTGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 157
Db 182 GCTCAGACCTGGGCGCCGAGATGCTTGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 241
Qy 158 ACGTGGCGGACTGGCTGGCGAGCAGATCAGGATCAGTCTGAAAAACAGCGTGA 217
Db 242 ACGTGGCGGAGCTGCTGGCGAGCAGATCAGGATCAGTCTGAAAAACAGCGTGA 301
Qy 218 TGAAGTGTGACGCGTGGCGGCGCAGCCG 246
Db 302 TGAAGTGTGACGCGTGGCGGATGACAGCAG 330

RESULT 14 381 bp mRNA linear EST 06-NOV-2001
LOCUS HOA37-1-f6 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
DEFINITION mRNA sequence.
ACCESSION BG900343
VERSION BG900343.1 GI:14310592
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 381)
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M., and
Lark,M.W.
Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthritic cartilage cDNA libraries
Osteoarthr. Cartil. 9 (7), 641-653 (2001)
JOURNAL MEDLINE 21482651
PUBMED 11597177
COMMENT Contact: Sanjay Kumar
UW2109
GlaxosmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-1@gsk.com
Seq primer: T7.
Location/Qualifiers
FEATURES
source 1. .381
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/clone_lib="HOA (Human Osteoarthritic Cartilage)"
/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 70 a 133 c 119 g 58 t 1 others
ORIGIN

Query Match 37.0%; Score 137.8; DB 12; Length 381;

Best Local Similarity 95.3%; Pred. No. 4.2e-22;
Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 98 GTTACAGCTGGGCGCCGAGATGCTTGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 157
Db 191 GCTCAGACCTGGGCGCCGAGATGCTTGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 250
Qy 158 ACGTGGCGGACTGGCTGGCGAGCAGATCAGGATCAGTCTGAAAAACAGCGTGA 217
Db 251 ACGTGGCGGAGCTGCTGGCGAGCAGATCAGGATCAGTCTGAAAAACAGCGTGA 310
Qy 218 TGAAGTGTGACGCGTGGCGGCGCAGCCG 246
Db 311 TGAAGTGTGACGCGTGGCGGATGACAGCAG 339

RESULT 15 383 bp mRNA linear EST 06-NOV-2001
LOCUS HOA8-1-F4 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
DEFINITION mRNA sequence.
ACCESSION BG989382
VERSION BG989382.1 GI:14308631
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 383)
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M., and
Lark,M.W.
Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthritic cartilage cDNA libraries
Osteoarthr. Cartil. 9 (7), 641-653 (2001)
JOURNAL MEDLINE 21482651
PUBMED 11597177
COMMENT Contact: Sanjay Kumar
UW2109
GlaxosmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-1@gsk.com
Seq primer: T7.
Location/Qualifiers
FEATURES
source 1. .383
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/clone_lib="HOA (Human Osteoarthritic Cartilage)"
/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 67 a 137 c 120 g 59 t
ORIGIN

Query Match 37.0%; Score 137.8; DB 12; Length 383;
Best Local Similarity 95.3%; Pred. No. 4.2e-22;
Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 98 GTTACAGCTGGGCGCCGAGATGCTTGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 157
Db 184 GCTCAGACCTGGGCGCCGAGATGCTTGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 243
Qy 158 ACGTGGCGGACTGGCTGGCGAGCAGATCAGGATCAGTCTGAAAAACAGCGTGA 217
Db 244 ACGTGGCGGAGCTGCTGGCGAGCAGATCAGGATCAGTCTGAAAAACAGCGTGA 303
Qy 218 TGAAGTGTGACGCGTGGCGGCGCAGCCG 246

Db 304 TGGAGTGTGACGCGGTGCGGGATGCACAG 332

Search completed: August 26, 2003, 15:34:29
Job time : 1434 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 26, 2003, 16:06:34 ; Search time 59 Seconds

(without alignments)
2001.569 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 672
Sequence: 1 aagcttaccatgggaagta.....aattccatcgcctcgcag 372

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=rlp
-Q=/cgnr_1/USPTO.spool/US09696872/runat_26082003_150541_8713/app_query.fasta_1.519
-DB=A-Geneseq.19Jun03 -QPM=faasan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pco -NOEM=ext -HEAPSIZE=500 -MIMLEN=0 -MAXLEN=2000000000
-USER=US09696872.QCGN_1.1117@runat_26082003_150541_8713 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUDRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A-Geneseq.19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	579	86.2	109	21	AA44963
2	459	68.3	109	21	AA44966
3	436.5	65.0	115	21	AA44959
4	429	63.8	109	21	AA44965
5	429	63.8	134	21	AA44967
6	422.5	62.9	115	21	AA44958
7	379.5	56.5	109	21	AA44962
8	353.5	52.6	109	21	AA44961
9	345.5	51.4	105	21	AA44960
10	304.5	45.3	90	21	AA44964
11	262.5	39.1	242	21	AA44964
12	262.5	39.1	300	21	AA44964
13	262.5	39.1	757	21	AA44964
14	262.5	39.1	757	21	AA44964
15	262.5	39.1	757	21	AA44964
16	262.5	39.1	757	21	AA44964
17	262.5	39.1	757	21	AA44964
18	256.5	38.2	130	22	AA44964
19	241	35.9	46	21	AA44964
20	211	31.4	46	21	AA44964
21	153	22.8	1357	22	AA44964
22	146.5	21.8	889	15	AA44964
23	146	21.7	35	15	AA44964
24	146	21.7	35	15	AA44964
25	146	21.7	54	15	AA44964
26	146	21.7	54	15	AA44964
27	145.5	21.7	776	22	AA44964
28	145.5	21.7	961	23	AA44964
29	145.5	21.7	961	23	AA44964
30	144	21.4	24	21	AA44964
31	144	21.4	24	21	AA44964
32	142.5	20.5	622	22	AA44964
33	138	20.5	74	18	AA44964
34	133	19.8	272	10	AA44964
35	129.5	19.3	163	23	AA44964
36	126.5	18.8	230	23	AA44964
37	126	18.8	329	23	AA44964
38	124	18.5	22	23	AA44964
39	123.5	18.4	1247	22	AA44964
40	123.5	18.4	1247	22	AA44964
41	123	18.3	814	22	AA44964
42	121	18.0	46	21	AA44964
43	121	18.0	667	19	AA44964
44	118	17.6	378	12	AA44964
45	117	17.4	46	21	AA44964

ALIGNMENTS

RESULT 1
ID AA44963 standard; Protein: 109 AA.
XX AA44963;
XX 23-MAY-2000 (first entry)
XX KDEL receptor inhibitor protein-6.
XX KDEL receptor inhibitor; heat shock protein; immune response;
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease.
XX Chimeric - Adenovirus E3.
XX Chimeric - Homo sapiens.
XX Chimeric - Camelus sp.

FH	Key Peptide	Location/Qualifiers 1..20 /label= Signal peptide
FT		/note= "Derived from adenovirus E3"
FT	Domain	30..75
FT		/note= "Human COMP pentamerisation domain"
FT	Domain	76..99
XX		/note= "Camel IgG linker domain"
PN		MO200006729-AI.
PD		10-FEB-2000.
PF		28-JUL-1999; 99WO-US17147.
XX		29-JUL-1998; 98US-0124671.
PR	(SLOK) SLOAN KETTERING INST CANCER RES.	
PA	Rothman JE, Mayhew M, Hoe MH;	
PI	WPI; 2000-195296/17.	
DR	N-PDB; AAZ50497.	
XX	Inhibitors of the KDEL receptor which comprises an oligomerization domain useful for promoting secretion of proteins which are normally retained within the cell -	
PT		
PS	Disclosure; Fig 6; 87bp; English.	
CC	The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteases by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several subunits where each subunit comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergy or autoimmune disease. The present sequence is KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; the oligomerisation domain of human cartilage oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequense GPDC is an alteration of rat COMP which provides increased stability via disulphide bonds.	
CC		
CC		
SQ	Sequence 109 AA:	
Alignment Scores:		
Pred. No.:	1,39e-47	Length: 109
Score:	579.00	Matches: 108
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	86.16%	Indels: 0
DB:	21	Gaps: 0
US-09-696-872-24 (1-372) x AAY44963 (1-109)		
OY	AGGTACATCATTTTAGCGTTCGCCCTTGCGGCAGACTGGAGCGCTGCCAAAAAGA	75
Dd	2 ArgTYMeLLleucnclYeuleuaLaLeuaLaLaValCYseerAlAblybuGeLy	21
OY	TCCAGCTGGGTGAGACTGTATTCAAGCTGGGCCCGCACAGATCTTCGGAACTGCAG	135
Dd	22 SetSerTeuGlVcJlYvPpcYCysSeSrAepLeugLYProGLmetLeuArgLUeuGln	41
OY	136 GAACAACCAGCGCGCGCTGCAGACGCTGCGGAGCTGCGTGGCAGACGATCGAGGAATC	195
Dd	42 GlutThrAsnaLaalaLeueunInspVaLAgaApTPLeuArvgLnGlnvalARgUlle	61

QY	196	ACGTTCTGAAAAACACGGTGAATGAGTGTGACGCGGTGCGGCGCACGCCGACCGGAAA	255
Db	62	ThreuleuylsanthrValmetGlucysaSPalaCysclProGlnProGlnProLys	81
QY	256	CCGACGCGCGGACGCGCGCGCGCGCGGAAACCGACGCGGAAACCGGAAAGGTATCC	315
Db	82	ProGlnProGlnProGlnProGlnProGlnProLysProGlnProLysProGlnProGlnIThr	101
QY	316	GGATCATCAGAAAAAGATGAGTTG	339
Db	102	GlySerSerGlyLysaSPaGlnLeu	109
RESULT 2			
ID	AAV44966	standard; Protein; 109 AA.	
AC	AAV44966		
XX	AAV44966;		
DT	23-MAY-2000	(first entry)	
XX			
DE	KDEL receptor inhibitor protein-9.		
XX			
KW	KDEL receptor inhibitor; heat shock protein; immune response;		
KW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;		
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;		
KW	infectious disease; allergy; autoimmune disease.		
XX			
OS	Chimeric - Adenovirus E3.		
OS	Chimeric - Homo sapiens.		
OS	Chimeric - Camelus sp.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..20	
FT		/label= "Signal peptide	
FT		/note= "Derived from adenovirus E3"	
FT	Domain	30..75	
FT		/note= "Human TSP4 trimerisation domain"	
FT	Domain	76..99	
FT		/note= "Camel IgG linker domain"	
PN	W0200006729-A1.		
XX			
PD	10-FEB-2000.		
XX			
PF	28-JUL-1999;	99WC-USL7147.	
XX			
PR	29-JUL-1998;	98US-0124671.	
XX			
PA	(SLOK) SLOAN KETTERING INST CANCER RES.		
XX			
PI	Rothman JE, Mayhew M, Hoe MH;		
DR	WPI; 2000-195296/17.		
XX	N-PSDB; AAZ50500.		
PT	Inhibitors of the KDEL receptor which comprises an oligomerization		
PT	domain useful for promoting secretion of proteins which are normally		
PT	retained within the cell		
XX			
PS	Disclosure; Fig 9; 87pp; English.		
XX			
CC	The patent discloses the use of KDEL receptor inhibitor to promote		
CC	secretion of proteins that are normally retained within the cell such as		
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of		
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat		
CC	shock proteins more accessible to the immune system and improves immune		
CC	response to a target antigen. The inhibitor protein comprises several		
CC	subunits where each subunit comprises an oligomerisation domain and has		
CC	at its carboxy terminus a region which binds to a KDEL receptor. The		
CC	target antigen may be associated with diseases including neoplasia such		
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and		
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,		

QY 274 CCGAGCCGAAACCCGACCCGAAACCCGAGTACCGATCATCGAAAAAGAT 333
 Db 94 ProGlnProIysProGlnProIysProGlnProIysGlyThrGlySerSerGluLysAsp 113
 QY 334 GAGTTG 339
 Db 114 GluLeu 115

RESULT 4
 ID AAY44965 standard; Protein; 109 AA.
 AC AAY44965;
 DT 23-MAY-2000 (first entry)
 DE KDEL receptor inhibitor protein-8.
 XX KDEL receptor inhibitor; heat shock protein; immune response;
 KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KW infectious disease; allergy; autoimmune disease.
 XX Chimeric - Adenovirus E3.
 OS Chimeric - Homo sapiens.
 OS Chimeric - Camelus sp.
 XX

Key Location/Qualifiers
 Peptide 1..20
 /label= Signal_peptide
 /note= "Derived from adenovirus E3"
 FT 30..75
 FT /note= "Human TSP 3 trimerisation domain"
 FT 76..99
 FT /note= "Camel IgG linker domain"
 XX

WO200006729-A1.
 PD 10-FEB-2000.
 XX
 PF 28-JUL-1999; 99WO-US17147.
 XX
 XX 29-JUL-1998; 98US-0124671.
 XX
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 PA
 XX Roehman JE, Mayhew M, Hoe MH;
 PI WPI; 2000-195296/17.
 DR N-PSDB; AAZ50499.
 XX

PT Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX

PS Disclosure; Fig 8, 87pp; English.
 XX

CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence is KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of human
 CC thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an

CC alteration of rat cartilage oligomeric matrix protein which provides
 CC increased stability via disulphide bonds.
 XX

Sequence 109 AA;
 SQ

Alignment Scores:
 Pred. No.: 4.24e-33 Length: 109
 Score: 429.00 Matches: 78
 Percent Similarity: 84.26% Conservative: 13
 Best Local Similarity: 72.22% Mismatches: 17
 Query Match: 63.84% Indels: 0
 DB: 21 Gaps: 0

US-09-696-872-24 (1-372) x AAY44965 (1-109)
 QY 16 AGGTACATGATTTTATGAGCTTCTCGCCCTTGCGGACGCTTCGACGCGCAAAAAGA 75
 Db 2 ArgTyrMetIleLeuGlyLeuLeuAlaLeuAlaValCysSerAlaIleValblybGly 21
 QY 76 TCCAGCCTGGGTGAGACTGTGTTTCAGACCTGGGCCCGCAGATGCTTCGGGAATGCGAG 135
 Db 22 SerSerLeuGlyGlyAspCyGcGlyGluGlnThrLysAlaLeuValThrGlnLeuThr 41
 QY 136 GAACCAACGCGCGCTGCGACGAGCTGCGGACTGCTGCGGACGAGTGAAGGATC 195
 Db 42 LeuPheAsnGlnIleLeuValGluLeuAlaArgAspIleArgAspGlnValblyGluMet 61
 QY 196 ACGTTCTGAAAAAACACGCGTATGAGTGAACCGCTGCGGCGCGACCGCAGCCGAAA 255
 Db 62 SerLeuIleArgAsnThrIleMetGluCyGlnValblybGlyProGlnProGlnProIys 81
 QY 256 CCGCAGCCGACGCGCGACCGCGACCCGAAACCGCAGCCGAAACCGGAACGGATACC 315
 Db 82 ProGlnProGlnProGlnProGlnProIysProGlnProIysProGlnProIysGlyThr 101
 QY 316 GGATCATCAGAAAAGATGAGTTG 339
 Db 102 GlySerSerGluLysAspGluLeu 109

RESULT 5
 ID AAY44967 standard; Protein; 134 AA.
 AC AAY44967;
 DT 23-MAY-2000 (first entry)
 DE KDEL receptor inhibitor protein-10.
 XX
 XX KDEL receptor inhibitor; heat shock protein; immune response;
 KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KW infectious disease; allergy; autoimmune disease.
 XX

OS Chimeric - Mus sp.
 OS Chimeric - Rattus sp.
 OS Chimeric - Camelus sp.
 XX

Key Location/Qualifiers
 Peptide 1..20
 /label= Signal_peptide
 /note= "Derived from mouse B1p"
 FT 23..32
 FT Region
 FT /label= Myc_tag_sequence
 FT 36..38
 FT Modified-site
 FT /note= "Asn is N-glycosylated"
 FT 49..94
 FT Domain
 FT /note= "Rat COMP pentamerisation domain"
 FT 101..124
 FT /note= "Camel IgG linker domain"
 XX

PN WO200006729-A1.

The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several subunits where each subunit comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergy or autoimmune diseases. The present sequence encodes KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; a myc-tag; an N-glycosylation sequence; the oligomerisation domain of rat cartilage oligomeric matrix protein (COMP); a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of rat COMP which provides increased stability via disulphide bonds.

US-09-696-872-24 (1-372) X AAY44967 (1-134)

[illegible]

XX	Sequence	115 AA:	
SQ	Alignment Scores:		
XX	Pred. No.:	1,81e-32	Length: 115
XX	Score:	422.50	Matches: 88
XX	Percent Similarity:	77.87%	Conservative: 7
XX	Best Local Similarity:	72.13%	Mismatches: 8
XX	Query Match:	62.87%	Indels: 19
DB:		21	Gaps: 4
US-09-696-872-24 (1-372) x AAY44958 (1-115)			
OY	10	ATGGAGAGTACATGATTTTA-----GGCTTGCTCGCCCTTGCGGAGCTGTGACGCGCT	63
DB	1	MetGlyLysPheThrValAlaAlaAlaLeuLeuLeuGlyAlaVal-----	17
OY	64	GCCAAAAAGATCCAGCTCGGTGGAGACTGTTTCAGACCTGGGCGCCGAGATGCTT	123
DB	18	ArgAlaGluGlySerSerLeuGlyGly-----AspLeuAlaProGlnMetLeu	33
OY	124	CGGGAATGTGCAAGAAACCAACCGCGCTGCAGACGTGCGGAGTGGCTGGCGAGCAG	183
DB	34	ArgGluLeuGlnGlnThrPheAlaAlaLeuGlnAspValArgGluLeuLeuArgGlnGln	53
OY	184	GTCAGGAGATCACGTTCTTGAAAAACAACGCTGATGAGTGTGACGCGTGGCGG-----	237
DB	54	ValLysGluIleThrPheLeuLysAsnThrValMetGluCysAspAlaCysGlyMetGln	73
OY	238	-----CCGCAGCCGACCGCGAAACCGCACCGCCGACCGCCGACG	273
DB	74	ProAlaArgThrProGlyThrSerProGlnProGlnProGlnProGlnProGln	93
OY	274	CCGCAAGCCGAAACCGCACCGCGAAACCGGAAACCGGAGGTACCGGATCATCGAAAAAGAT	333
DB	94	ProGlnProLysProGlnProGlnProLysProGlnProGlnGlyThrGlySerSerGluLysAsp	113
OY	334	GAGTTG 339	
DB	114	GluLeu 115	
RESULT 7			
AAY44962			
ID	AAY44962 standard; Protein; 109 AA.		
XX	AAY44962;		
XX	AC		
XX	23-MAY-2000 (first entry)		
XX	XX		
XX	KDEL receptor inhibitor protein-5.		
XX	XX		
KW	KDEL receptor inhibitor; heat shock protein; immune response;		
KW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;		
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;		
KW	infectious disease; allergy; autoimmune disease.		
XX	XX		
OS	Chimeric - Mus sp.		
OS	Chimeric - Xenopus sp.		
OS	Chimeric - Camelus sp.		
OS	Chimeric - Rattus sp.		
XX	XX		
Key	Location/Qualifiers		
FT	Peptide	1..20	
FT	/label=	Signal peptide	
FT	/notes=	Derived from mouse Bfp"	
FT	Domain	26..30	
FT	/notes=	"Altered subsequence from rat COMP"	
FT	Domain	30..75	
FT	/notes=	"Xenopus thrombospondin 4 trimerisation domain"	
FT	Domain	76..99	
FT	/notes=	"Camel IgG linker domain"	
XX	XX		
PN	W0200006729-A1.		

XX	XX		
PD	10-FEB-2000.		
XX	XX		
PF	28-JUL-1999;	99WO-US17147.	
XX	XX		
PR	29-JUL-1998;	98US-0124671.	
XX	XX		
PA	(SLOK) SLOAN KETTERING INST CANCER RES.		
XX	XX		
PI	Rothman JE, Mayhew M, Hoe MH;		
XX	XX		
DR	WPI; 2000-195296/17.		
DR	N-PSDB; AAZ50496.		
PT	Inhibitors of the KDEL receptor which comprises an oligomerization		
PT	domain useful for promoting secretion of proteins which are normally		
PT	retained within the cell		
XX	XX		
PS	Disclosure; Fig 5; 87pp; English.		
XX	XX		
CC	The patent discloses the use of KDEL receptor inhibitor to promote		
CC	secretion of proteins that are normally retained within the cell such as		
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of		
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat		
CC	shock proteins more accessible to the immune system and improves immune		
CC	response to a target antigen. The inhibitor protein comprises several		
CC	subunits where each subunit comprises an oligomerisation domain and has		
CC	at its carboxy terminus a region which binds to a KDEL receptor. The		
CC	target antigen may be associated with diseases including neoplasia such		
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and		
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,		
CC	infectious diseases, allergy or autoimmune diseases. The present		
CC	sequence is KDEL receptor inhibitor comprising regions encoding a		
CC	cleavable signal peptide; the oligomerisation domain of Xenopus		
CC	thrombospondin 4 (TSP4) trimerisation domain including an additional		
CC	subsequence; a camel IgG linker domain and the carboxy-terminal sequence		
CC	KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric		
CC	matrix protein which provides increased stability via disulphide bonds.		
XX	XX		
SQ	Sequence 109 AA;		
Alignment Scores:			
Pred. No.:	2.55e-28	Length: 109	
Score:	379.50	Matches: 70	
Percent Similarity:	77.68%	Conservative: 17	
Best Local Similarity:	62.50%	Mismatches: 20	
Query Match:	56.47%	Indels: 5	
DB:	21	Gaps: 2	
US-09-696-872-24 (1-372) x AAY44962 (1-109)			
OY	10	ATGGAGAGTACATGATTTTA-----GGCTTGCTCGCCCTTGCGGAGCTGTGACGCGCT	63
DB	1	MetGlyLysPheThrValAlaAlaAlaLeuLeuLeuGlyAlaVal-----	17
OY	64	GCCAAAAAGATCCAGCTCGGTGGAGACTGTTTCAGACCTGGGCGCCGAGATGCTT	123
DB	18	ArgAlaGluGlySerSerLeuGlyGlyAspCysCysGlyAspValSerArgGlnLeuIle	37
OY	124	CGGGAATGTGCAAGAAACCAACCGCGCTGCAGACGTGCGGAGCTGCTGGCGAGCAG	183
DB	38	GlyGlnIleThrGlnMetAsnGlnMetLeuGlyGluLeuArgAspValMetArgGlnGln	57
OY	184	GTCAGGAGATCACGTTCTTGAAAAACAACGCTGATGAGTGTGACGCGTGGCGGCGCAG	243
DB	58	ValLysGluThrMetPheLeuArgAsnThrIleAlaGluCysGlnAlaCysGlyProGln	77
OY	244	CCGCAAGCCGAAACCGCACCGCGAGCCGCAACCGGAAACCGCAACCGGAAACCGGAA	303
DB	78	ProGlnProLysProGlnProGlnProGlnProLysProGlnProLysProGlnProGln	97
OY	304	CCGGAAGTACCGGATCATGAAAAAGATGAGTTG 339	

Alignment Scores:			
Pred. No.:	8, 266-26	Length:	109
Score:	353.50	Matches:	66
Percent Similarity:	75.00%	Conservative:	18
Best Local Similarity:	58.93%	Mismatches:	23
Query Match:	52.60%	Indels:	5
DB:	21	Gaps:	2
US-09-696-872-24 (1-372) x AAY44961 (1-109)			
QY	10 ATGGAGAGTACATGATTTTA-----GACTTGCTCGCCCTTGCGGACGTGACGCGCT	63	
Db	1 MetcLylsePhehthrValValAlaAlaAlaLeuLeuLeuGlyAlaVal-----	17	
QY	64 GCCAAMAAAAGATCCAGCGCTGGGTGAGACTGTTGTCAGACTGGGCCCGCAATGCTT	122	
Db	18 ArgAlaGlucllySerSerLeucllyGlyAspCysCysGlyGluGlnThrLysAlaLeuVal	37	
QY	124 CGGAACTGACAGAAACCAACCGCGCGCTGCAGCGAGCGGAGACTGGCTCGGCAGAG	181	
Db	38 ThGlnLeuThrIeuPheAsnGlnIleLeuValGluLeuAspAspIleArgAspGln	57	
QY	184 GTCCAGAGATCACTGTTCTCTGAAAAACAAGGTATGATGAGTGTGACGCGTGGCGCCGAG	243	
Db	58 VallyGluMeSerLeuIleArgAsnThrLysMetGluCysGlnValCysGlyProGln	77	
QY	244 CGCGACCCGAAACCGGAGCCGACCGCGACCGCGGAAACCGCACCGAAACCGGAA	302	
Db	78 ProGlnProlyseProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln	97	
QY	304 CGGAAAGTACCGGATCATCAGAAAAAGATGAGTGTG	339	
Db	98 ProGluGlyThrGlySerSerGluLysAspGlnLeu	109	
RESULT 9			
AAV44960	standard; Protein; 105 AA.		
AC	AAY44960;		
XX			
XX	23-MAY-2000 (first entry)		
XX			
XX	KDEL receptor inhibitor protein-3.		
DE			
XX	KDEL receptor inhibitor; heat shock protein; immune response;		
KW	K oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;		
KM	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;		
KW	infectious disease; allergy; autoimmune disease.		
XX			
OS	Chimeric - Mus sp.		
OS	Chimeric - Camelus sp.		
OS	Chimeric - Rattus sp.		
XX			
XX	Key		
FT	Peptide	Location/Qualifiers	
FT		1..20	
FT		/label= Signal peptide	
FT		/note= "Derived from mouse BIP"	
FT	Domain	26..30	
FT		/note= "Altered subsequence from rat COMP"	
FT	Domain	30..71	
FT		/note= "Mouse thymospondin 3 trimerisation domain"	
FT	Domain	72..95	
FT		/note= "Camel IgG linker domain"	
XX	WO200006729-A1.		
XX			
XX	10-FEB-2000.		
XX			
XX	28-JUL-1999; 99WO-US17147.		
XX			
XX	29-JUL-1998; 98US-0124671.		
XX			

PA (SLOK) SLOAN KETTERING INST CANCER RES.
 XX Rothman JE, Mayhew M, Hoe MH;
 XX MPI: 2000-195296/17.
 DR N-PSDB; AAZ50494.
 XX
 XX Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX
 XX Disclosure: Fig 3; 87pp; English.
 XX
 CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerization domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence is KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of mouse
 CC thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an
 CC alteration of rat cartilage oligomeric matrix protein which provides
 CC increased stability via disulphide bonds.
 XX
 XX Sequence 105 AA;
 SQ
 Alignment Scores:
 Pred. No.: 4,86e-25 Length: 105
 Score: 345.50 Matches: 67
 Percent Similarity: 74.11% Conservative: 16
 Best Local Similarity: 59.82% Mismatches: 20
 Query Match: 51.41% Indels: 9
 DB: 21 Gaps: 3
 US-09-696-872-24 (1-372) x AAY44960 (1-105)
 QY 10 ATGGAGAGTACATGATTTTA-----GGCTTGTCGCGCTTGGCAGTGTGACGCCT 63
 DB 1 MetGlyLysPheThrValValAlaAlaLeuLeuLeuGlyAlaVal----- 17
 QY 64 GCCAAAAAAGATCCAGCCTGGGTGAAGACTGTGTTACAGACTGGGCCCGCAGATGCTT 123
 DB 18 ArgAlaGluGlySerSerLeuGlyLysPyrCybLysAlaLeu-----Val 33
 QY 124 CGGGAACTCGACGGAACCAACCGCGCGCTGCGACGACCTGCGGACTGGCTGGCAGCAG 183
 DB 34 ThrGlnLeuThrLeuPheAsnGlnIleLeuValGluLeuAlaAspAspIleArgAspGln 53
 QY 184 GTTCAGGAGATCACCTCTCTGTAACCAACGAGTGAAGTGAAGCGCTGGCGCGCAGAG 243
 DB 54 VallysGluMetSerLeuLeuLeuArgAsnThrIleMetGluCybGlnValCysGlyProGln 73
 QY 244 CGGACCGCGGAACCGCAGCGCGACCGCGACCGCGAACCAGCCGGAACCGGAACCGGA 303
 DB 74 ProGlnProLysPyrProGlnProGlnProGlnProGlnProLysPyrProGlnProGlu 93
 QY 304 CCGGAAGGTACCGGATCATCGAAGAAAGATGAGTGG 339
 DB 94 ProGlnGlyThrGlySerSerGlnLysAspGlnLeu 105
 RESULT 10
 ID AAY44964 standard; Protein; 90 AA.
 XX AAY44964;
 AC AAY44964;

XX 23-MAY-2000 (first entry)
 DT
 XX KDEL receptor inhibitor protein-7.
 DE
 XX KDEL receptor inhibitor; heat shock protein; immune response;
 XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
 KW infectious disease; allergy; autoimmune disease.
 XX
 OS Chimeric - Adenovirus E3.
 OS Chimeric - Homo sapiens.
 OS Chimeric - Camelus sp.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..20
 FT /note= "Derived from adenovirus E3"
 FT Domain 30..56
 FT /note= "Human phospholamban pentamerisation domain"
 FT Domain 57..80
 FT /note= "Camel IgG linker domain"
 XX
 XX WO200006729-A1.
 XX
 XX 10-FEB-2000.
 XX
 XX 28-JUL-1999; 99WO-US17147.
 XX
 XX 29-JUL-1998; 98US-0124671.
 XX
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 PA
 XX Rothman JE, Mayhew M, Hoe MH;
 XX MPI: 2000-195296/17.
 DR N-PSDB; AAZ50498.
 XX
 XX Inhibitors of the KDEL receptor which comprises an oligomerization
 PT domain useful for promoting secretion of proteins which are normally
 PT retained within the cell -
 XX
 XX Disclosure: Fig 7; 87pp; English.
 PS
 XX
 CC The patent discloses the use of KDEL receptor inhibitor to promote
 CC secretion of proteins that are normally retained within the cell such as
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat
 CC shock proteins more accessible to the immune system and improves immune
 CC response to a target antigen. The inhibitor protein comprises several
 CC subunits where each subunit comprises an oligomerisation domain and has
 CC at its carboxy terminus a region which binds to a KDEL receptor. The
 CC target antigen may be associated with diseases including neoplasia such
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,
 CC infectious diseases, allergy or autoimmune diseases. The present
 CC sequence is KDEL receptor inhibitor comprising regions encoding a
 CC cleavable signal peptide; the oligomerisation domain of human
 CC phospholamban (PLB) pentamerisation domain; a camel IgG linker domain
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an
 CC alteration of rat cartilage oligomeric matrix protein which provides
 CC increased stability via disulphide bonds.
 XX
 XX Sequence 90 AA;
 SQ
 Alignment Scores:
 Pred. No.: 4.32e-21 Length: 90
 Score: 304.50 Matches: 67
 Percent Similarity: 60.18% Conservative: 1
 Best Local Similarity: 59.29% Mismatches: 16
 Query Match: 45.31% Indels: 29
 DB: 21 Gaps: 2
 US-09-696-872-24 (1-372) x AAY44964 (1-90)

QY 16 AGGTACATGATTATTAAGGCTTCTGCGCCCTTCGCGCAGTCTGACGCGTGCACCAAAAAGA 75
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 Db 2 Arglyrmetileuleuglyleuleualaleuualaleuualaleuualaleuualaleuualaleu 21
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 QY 76 TCCAGCCTGGGTGAGACTGTTGTTACAGACTGGGCGCCGACAGATGCTTGGGAACTGCAG 135
 CC Sequence 242 AA:
 Db 22 SerseirleuglyglyAspCysCysglInlyseu----- 32
 QY 136 GAACCAACGCGGGGCTGCAGAGCGTGGGAGACTGGCTGCGCGACGATCAGGAGATC 195
 Db 33 -----GlnAenleuphelle 37
 QY 196 ACGTTCCTGAAAAACACGCGTATGATGATGACCGTGC-----GCGCCG 240
 Db 38 Asnhecysleuileuleuileuileuileuileuileuileuileuileuileuileuileu 57
 QY 241 CAGCGCGACCGGAAACCGCAGCGCGCAGCGCGCGGAAACCGCAGCGCGGAAACCG 300
 Db 58 GlnProGlnProLysProGlnProGlnProGlnProGlnProGlnProGlnProLysPro 77
 QY 301 GAACCGAAGGTACCGATCATCATGAAAAAGATGATTG 339
 Db 78 GlnProGlnGlyThrGlySerSerGlnLysAspGlnLeu 90
 RESULT 11
 AAB00040 ID AAB00040 standard; Protein: 242 AA.
 AC AAB00040;
 DT 08-NOV-2000 (first entry)
 DE Human COMP/TSP-1 chimeric protein.
 XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glaucoma.
 OS Homo sapiens.
 OS Synthetic.
 XX WO200044908-A2.
 PN 03-AUG-2000.
 PD 01-FEB-2000; 2000WO-US02482.
 PF 01-FEB-1999; 99US-0118053.
 PR 01-FEB-1999; 99US-0118053.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA Lawler JW;
 PI WPI: 2000-514823/46.
 DR N-PSDB; AAA47734.
 PT Nucleic acid encoding chimeric proteins such as cartilage oligomeric
 PT matrix protein (COMP)/thrombospondin (TSP)-1 and 2, useful for
 PT inhibiting angiogenesis and treating diseases such as cancer
 XX Claim 33; Fig 4a-b; 40pp; English.
 PS New nucleic acids are described which encode a protein comprising
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,
 CC but not the TGF (transforming growth factor)-beta activation region
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing
 CC the second and third type-1 repeats and the COMP (cartilage
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1
 CC caused inhibition of the growth of tumours in mice models.
 CC Thus the nucleic acids and proteins may be useful for treating

CC angiogenesis related diseases such as cancer (by reducing the rate of
 CC growth and size of tumours), arthritis, psoriasis, diabetic
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be
 CC used for treating human immunodeficiency virus (HIV) infection.
 CC Anti-angiogenic therapy has little toxicity, does not require the
 CC therapeutic agent to enter tumour cells or cross the blood-brain
 CC barrier, controls tumour growth independently of growth of
 CC tumour cell heterogeneity, and does not induce drug resistance.
 XX
 SO Sequence 242 AA:
 Alignment Scores:
 Pred. No.: 5,73e-17 Length: 242
 Score: 262.50 Matches: 57
 Percent Similarity: 75.61% Conservative: 5
 Best Local Similarity: 69.51% Mismatches: 9
 Query Match: 39.06% Indels: 11
 DB: 21 Gaps: 2
 US-09-696-872-24 (1-372) x AAB00040 (1-242)
 QY 34 TTGTCGCCCTTGGCGGAGTCTGACGCGTGCACCAAAAAGATCCAGCTGGGTGAGAC 93
 Db 10 LeuLeuThrleuAlaAlaLeuGlyAlaSerGlyGlnGlyGlnSerProleuGly----- 27
 QY 94 TGTTCAGACCTGGGCGCGCAGATGCTTGGGAACTGCAGGAAACCAACGCGCGCTG 153
 Db 28 -----SerAspLeuGlyProGlnMetLeuArgGlnLeuGlnThrAsnAlaAlaLeu 45
 QY 154 CAGGACGTGGGGAAGTGGCTGCGCGCAGCAGGTGAGGAGATCATCGTTCGTAACACG 213
 Db 46 GlnAspValArgAspTrpLeuArgGlnGlnValArgGlnIleThrPheLeuLysAsnThr 65
 QY 214 GTATGAGTGTGACCGGTGCGGCGCGCGAG-----CGGACCG 252
 Db 66 ValMetLeuLysAspAlaCysGlyMetGlnGlnSerValArgThrGlyLeuProSerVal 85
 QY 253 AAACCG 258
 Db 86 ArgPro 87
 RESULT 12
 AAB00041 ID AAB00041 standard; Protein: 300 AA.
 AC AAB00041;
 DT 08-NOV-2000 (first entry)
 DE Human COMP/TSP-2 chimeric protein.
 XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;
 KW glaucoma.
 OS Homo sapiens.
 OS Synthetic.
 XX WO200044908-A2.
 PN 03-AUG-2000.
 PD 01-FEB-2000; 2000WO-US02482.
 PF 01-FEB-1999; 99US-0118053.
 PR 01-FEB-1999; 99US-0118053.
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA Lawler JW;
 PI WPI: 2000-514823/46.
 DR N-PSDB; AAA47735.

QY 214 GTGATGAGTGTGACGGCTGCGGCGCCGAG-----CCGACGCCG 252
DB 66 ValmetglucysaapalacysgIymetGInglInserValArgThrGlyLeuProSerVal 85
QY 253 AAACCG 258
DB 86 ArgPro 87
RESULT 14
ABU05594
ID ABU05594 standard; Protein; 757 AA.
XX AC ABU05594;
XX 14-NOV-2002 (first entry)
XX DB Breast cancer-associated protein 59.
XX DE Breast cancer-associated protein 59.
XX KW Breast cancer; breast cancer-associated gene sequence;
XX KM drug development; pharmacogenetics; biosensor development.
XX OS Unidentified.
XX WO200259377-A2.
XX PD 01-AUG-2002.
XX PF 24-JAN-2002; 2002WO-US02242.
XX PR 24-JAN-2001; 2001US-263965P.
XX PR 02-FEB-2001; 2001US-265928P.
XX PR 09-APR-2001; 2001US-282698P.
XX PR 09-APR-2001; 2001US-0829472.
XX PR 04-MAY-2001; 2001US-288590P.
XX PR 29-MAY-2001; 2001US-294443P.
XX PA (BOSB-) EOS BIOTECHNOLOGY INC.
XX PI Mack DH, Gish KC, Afar D;
XX DR WPI; 2002-583738/62.
XX DR N-PSDB; ABT07751.
XX PT Detecting a breast cancer-associated transcript in a patient's cell,
XX PT useful for diagnosing breast cancer, comprises contacting a biological
XX PT sample with a polynucleotide that selectively hybridizes with breast
XX PT cancer nucleic acids
XX PS Disclosure; Page 401; 414pp; English.
XX CC The invention comprises a method of detecting a breast cancer-associated
XX CC transcript in a cell from a patient. The method of the invention involves
XX CC contacting a biological sample from the patient with a nucleotide that
XX CC hybridizes to one of the 69 breast cancer-associated gene sequences shown
XX CC in the specification. The method of the invention is useful in the
XX CC diagnosis or prognosis of breast cancer, and for detecting genes that are
XX CC up or down-regulated in breast cancer cells. Genes identified by the
XX CC method of the invention can be used in diagnostic purposes and also as
XX CC targets for screening for therapeutic compounds that modulate breast
XX CC cancer (e.g. hormones or antibodies). Identification of genes that are
XX CC over or under expressed in breast cancer can additionally provide high-
XX CC resolution, high-sensitivity datasets which can be used in the areas of
XX CC diagnostics, therapeutics, drug development, pharmacogenetics, protein
XX CC structure and biosensor development. Amino acid sequences ABU0536 -
XX CC ABU05604 represent the proteins encoded by the 69 breast cancer-
XX CC associated genes of the invention.
XX SQ Sequence 757 AA;

Alignment Scores:
Pred. No.: 6,85e-17 Length: 757
Score: 262.50 Matches: 57
Percent Similarity: 75.61% Conservative: 5

Best Local Similarity: 69.51% Mismatches: 9
Query Match: 39.06% Indels: 11
DB: 23 Gaps: 2
US-09-696-872-24 (1-372) x ABU05594 (1-757)
QY 34 TTGCTGCGCCCTTGGCGAGTCTGCGAGCGCTGCCAAAAAGATCCAGCTGGTGGAGAC 93
DB 10 LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyGlnGlyGlnSerProLeuGly----- 27
QY 94 TGTTCGACAGCTCGGAGCCCGGACATGCTTCGGGAACTGCAGGAAACCAACGCGCGCTG 153
DB 28 -----SerAspLeuGlyProGlnMetLeuArgGluLeuGlnGlnThrAsnAlaAlaLeu 45
QY 154 CAGGACGTGCGGAGCTGCGCTGCGGACAGAGTCAAGGAGATCAAGTCTGAAAAACAGC 213
DB 46 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuThrPheLeuValAsnThr 65
QY 214 GTGATGAGTGTGACGGCTGCGGCGCCGAG-----CCGACGCCG 252
DB 66 ValmetglucysaapalacysgIymetGInglInserValArgThrGlyLeuProSerVal 85
QY 253 AAACCG 258
DB 86 ArgPro 87
RESULT 15
ABR47420
ID ABR47420 standard; Protein; 757 AA.
XX AC ABR47420;
XX DT 12-JUN-2003 (first entry)
XX DE Breast cancer associated protein sequence SEQ ID NO:72.
XX DE Human; breast cancer; cytostatic; gene therapy.
XX KW Homo sapiens.
XX OS
XX KM Homo sapiens.
XX WO2003004989-A2.
XX PD 16-JAN-2003.
XX PF 21-JUN-2002; 2002WO-US19669.
XX PR 21-JUN-2001; 2001US-299887P.
XX PR 27-JUN-2001; 2001US-301572P.
XX PR 18-JUL-2001; 2001US-306501P.
XX PR 25-SEP-2001; 2001US-325002P.
XX PR 05-MAR-2002; 2002US-362585P.
XX PR 14-MAY-2002; 2002US-380391P.
XX PA (MILL-) MILLENIUM PHARM INC.
XX PI Lillie J, Gannavarapu M, Glatc K, Hoerish S, Kamalakar S, Mertens M,
XX PI Monahan JB, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Baer RC;
XX PI Horrobaght GW, Fuzestai L, Meric F, Sahin A, Mills GB;
XX DR WPI; 2003-210381/20.
XX DR N-PSDB; ACC50112.
XX PT Breast cancer diagnosis or treatment by comparing the level of
XX PT expression of a marker in a patient sample with that in the control
XX PT non-breast cancer sample
XX PS Claim 1; SEQ ID 72; 128pp; English.
XX CC The present invention describes a method for assessing whether a patient
XX CC is afflicted with breast cancer. The method comprises comparing the level
XX CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
XX CC ABR47386 to ABR47632) in a patient sample and the normal level of
XX CC expression of the marker in a control non-breast cancer sample, where a

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 26, 2003, 16:19:24 ; Search time 15.5 Seconds
(without alignments)
2030.921 Million cell updates/sec

Title: US-09-696-872-24
Perfect score: 672
Sequence: 1 aagcttcacatgggaagta.....aatccatcatgcctcgcag 372

Scoring table: BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues
Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p,model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09696872/runat_26082003_150544_8745/app_query.fasta.1.519
-DB=Issued_Patents_AA -QMT=faстан -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NOM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USBR=US09696872.qcgn_1.1_38.@runat_26082003_150544_8745 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCtus.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfllea1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	579	86.2	109	3	US-09-124-671-23
2	459	68.3	109	3	US-09-124-671-29
3	436.5	65.0	115	3	US-09-124-671-15
4	429	63.8	109	3	US-09-124-671-27
5	422.5	62.9	115	3	US-09-124-671-34
6	379.5	56.5	109	3	US-09-124-671-13
7	379.5	56.5	109	3	US-09-124-671-21
8	353.5	51.4	109	3	US-09-124-671-19
9	345.5	45.3	103	3	US-09-124-671-17
10	304.5	45.3	90	3	US-09-124-671-25
11	241	35.9	46	3	US-09-124-671-2
12	234.5	34.9	66	3	US-09-091-814-46

13	228	33.9	64	3	US-09-091-814-45	Sequence 45, Appl
14	211	31.4	46	3	US-09-124-671-1	Sequence 1, Appl1
15	146.5	21.8	889	5	PCT-US93-11725-2	Sequence 2, Appl1
16	146	21.7	35	1	US-08-471-780C-38	Sequence 38, Appl
17	146	21.7	35	1	US-08-467-282B-38	Sequence 38, Appl
18	146	21.7	35	2	US-08-471-282A-38	Sequence 38, Appl
19	146	21.7	35	2	US-08-466-710C-38	Sequence 38, Appl
20	146	21.7	35	3	US-08-468-739C-38	Sequence 38, Appl
21	146	21.7	54	1	US-08-471-780C-44	Sequence 44, Appl
22	146	21.7	54	1	US-08-467-282B-44	Sequence 44, Appl
23	146	21.7	54	2	US-08-471-282A-44	Sequence 44, Appl
24	146	21.7	54	2	US-08-466-710C-44	Sequence 44, Appl
25	146	21.7	54	3	US-08-468-739C-44	Sequence 44, Appl
26	146	21.7	60	1	US-08-471-780C-87	Sequence 87, Appl
27	146	21.7	60	1	US-08-467-282B-87	Sequence 87, Appl
28	146	21.7	60	2	US-08-471-282A-87	Sequence 87, Appl
29	146	21.7	60	2	US-08-466-710C-87	Sequence 87, Appl
30	146	21.7	60	3	US-08-468-739C-87	Sequence 87, Appl
31	145.5	21.7	961	5	PCT-US93-11725-4	Sequence 4, Appl1
32	144	21.4	24	4	US-09-379-297-5	Sequence 5, Appl1
33	121	18.0	46	3	US-09-124-671-5	Sequence 5, Appl1
34	121	18.0	667	2	US-08-718-661-2	Sequence 2, Appl1
35	119	17.0	271	4	US-09-252-991A-17292	Sequence 17292, A
36	117	17.4	46	3	US-09-124-671-6	Sequence 6, Appl1
37	116	17.3	261	4	US-09-602-565-34	Sequence 34, Appl
38	113.5	16.9	221	4	US-09-069-023-22	Sequence 22, Appl
39	112	16.7	907	3	US-08-989-399-12	Sequence 12, Appl
40	111	16.5	679	4	US-09-252-991A-18857	Sequence 18857, A
41	110	15.7	2211	4	US-09-738-884-1	Sequence 1, Appl1
42	109	16.2	558	4	US-09-252-991A-17202	Sequence 17202, A
43	108	16.1	304	4	US-09-252-991A-23116	Sequence 23116, A
44	107	15.9	447	4	US-09-252-991A-22113	Sequence 22113, A
45	107	15.9	613	4	US-09-345-473E-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-09-124-671-23
; Sequence 23, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PASTESEQ for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: chimeric human COMP-KDEL
US-09-124-671-23
Alignment Scores:
Pred. No.: 2.3e-50
Score: 579.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 86.16%
Matches: 109
Conservative: 108
Mismatch: 0
Indels: 0
Gaps: 0
US-09-696-872-24 (1-372) x US-09-124-671-23 (1-109)
QY 16 AGTACATGATTTTGGCTTGGCCCTTGCGGACGCTGCGCAAAAAGCA 75
DB 2 ArgYrmettleleuglyLeuvalaleuValaValaCysSerAlaAlaYslyGly 21


```

QY 10 ATGGAAGGATCAATGATTTT-----GGCTTGCCTGCCTTGCAGAGTGTGCAGCGCT 63
      ||||| :||| |||||
Db 1 MetGlyLysPheThrValValAlaAlaAlaLeuLeuLeuGlyAlaVal----- 17
QY 64 GCCAAAGATTCACGCGCTGGGTGAGAGCTTGTTCAGACCTGGGCGCGCAGATGCTT 122
      :||| :||| :||| :|||
Db 18 ArgAlaGlnGlySerLeuGlnGlyLys-----AspLeuAlaProGlnMetLeu 33

```

Qy	121	CGGGAAGTCCAGGAAACCAACCGCGCGCTCGACGACGTCGCGGACTGCTCGCGACACAG	183
Db	34	ArgGluLeuGlnGluThrAsnAlaIleuGlnAspValArgGluLeuLeuArgGlnGln	53
Qy	184	GTCCAGGAGATCACTGTTCTTGAAAAACAACGGATGATGATGATGACGCGTGGCGG-----	237
Db	54	ValLyGluIleThrPheLeuYsAsnThrValMetGluCybAphA1CybGlyMetGln	73
Qy	238	-----CCGACGCCGACCGGAAACCGACGCGACGCGCGACG	273
Db	74	ProAlaArgThrProGlyThrSerProGlnProGlnProPolysProGlnProGln	93
Qy	274	CCGACCGGAAACCGGACCGGAAACCGGAAACCGGAAAGGTACCGGATCATCGAAAAAGAT	333
Db	94	ProGlnProPolysProGlnProPolysProGlnProGlnGlyThrGlySerSerGlyAsp	113
Qy	334	GAGTTG 339	
Db	114	GluLeu 115	
RESULT 7			
US-09-124-671-21			
; Sequence 21, Application US/09124671A			
; Patent No. 6160088			
; GENERAL INFORMATION:			
; APPLICANT: Rothman, James			
; APPLICANT: Mayhew, Mark			
; APPLICANT: Hoe, Mee			
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS			
; FILE REFERENCE: 31488			
; CURRENT APPLICATION NUMBER: US/09/124,671A			
; CURRENT FILING DATE: 1998-07-29			
; NUMBER OF SEQ ID NOS: 42			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 21			
; LENGTH: 109			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: chimeric Xenopus laevis TSP4-KDEL			
US-09-124-671-21			
Alignment Scores:			
Pred. No.: 2.57e-30 Length: 109			
Score: 379.50 Matches: 70			
Percent Similarity: 77.68% Conservative: 17			
Best Local Similarity: 62.50% Mismatches: 20			
Query Match: 56.47% Indels: 5			
DB: 3 Gaps: 2			
US-09-696-872-24 (1-372) x US-09-124-671-21 (1-109)			
Qy	10	ATCGGAAGGTACATGATTTTA-----GGCTTGCTCGCGCTTGCGGAGTCTGCAGGCGCT	63
Db	1	MeGlyLyvPheThrValValAlaIleuLeuLeuGlyAlaVal-----	17
Qy	64	GCCAANAAGATTCAGCCTCGGTGGAGACTGTTGTCAGACTCGGCGCGAGATGCTT	123
Db	18	ArgAlaGluGlySerSerIleuGlyIleYsPcySeyGlyAspValIserArgGlnIleIle	37
Qy	124	CGGAACTGCAGAAACCAACCGCGCGCTGCAGACGTCGCGGACTGCTGCGGACAG	183
Db	38	GlyIleIleThrGlnMetAsnGlnMetIleuGlyIleuArgAspValMetArgGlnGln	57
Qy	184	GTCCAGGAGATCACTGTTCTTGAAAAACAACGATGATGATGATGACGCGTGGCGGCGCAG	243
Db	58	ValLyGluIleThrPheLeuYsAsnThrIleAlaGluCybGlnAlaCybGlyProGln	77
Qy	244	CCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGAAACCGGAAACCGGAA	303
Db	78	ProGlnProPolysProGlnProGlnProGlnProGlnProGlnProPolysProGlnProGln	97
Qy	304	CCGGAAGTACCGGATCATCAGAAAAAGATGAGTTG 339	

```

Db      98  ProGluGlyThrIysSerSerGluIuysAspGluLeu 109
|||||
RESULT 8
US-09-124-671-19
; Sequence 19, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric mouse TSP3-KDEL
US-09-124-671-19

Alignment Scores:
Pred. No.: 1,05e-27 Length: 109
Score: 353.50 Matches: 66
Percent Similarity: 75.00% Conservative: 18
Best Local Similarity: 58.93% Mismatches: 23
Query Match: 52.60% Indels: 5
DB: 3 Gaps: 2

US-09-696-872-24 (1-372) x US-09-124-671-19 (1-109)
QY      10  ATGGGAAGTACATGATTTT-----GGCTTGCTGCGCCTTGCGGACGTGACGCGCT 63
|||||
Db      1  MetGILyIyepheThrValAlaAlaAlaLeuLeuLeuGlyAlaVal----- 17
|||||

QY      64  GCCAAAGATCCAGCCCTGGGTGGAGACGTGTTGCACACCTGGGCGGCAGATGCTT 122
|||||
Db      18  ArgAlaGluGlySerSerIeuLeuIyAlaSpCySGlyGluGlnTrIyAlaLeuVal 37
|||||

QY      124 CGGAACTGCAGAAACCAACCGCGCGCTGCAGACGTGCGGACTGGCTGCGGACAG 183
|||||
Db      38  ThrGlnIeuThrIeuPhehngInIleuValAlGluLeuAlaArgAspAlaIeArgAspGln 57
|||||

QY      184 GTACAGGAATCACGTTCTGAAAAACACGCGTGAATGAGGTGACGCGTGGCGGCCAG 243
|||||
Db      58  ValIyGluMetSerIeuLeuIleArgAsnThrIleMetGluCySGlnValCySGlyProGln 77
|||||

QY      244 CGCGACCGCAACCGGAGCGGACGCGGACGCGGACGCGCAACCGGACCGCAACCGGAA 303
|||||
Db      78  ProGlnIeuIySProGlnIProGlnIProGlnIProGlnIProGlnIProGlnIProGln 97
|||||

QY      304 CCGAAGGTACCGGATCATCAGAAAAAGATGATTG 339
|||||
Db      98  ProGluGlyThrGlySerSerGluIuysAspGluLeu 109
|||||

RESULT 9
US-09-124-671-17
; Sequence 17, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0

```



```

? SEQ ID NO: 17
? LENGTH: 105
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chimeric mouse
? OS-09-124-671-17
? TSF3-KDEL

```

Alignment Scores:	
Pred. No.:	6,66e-27
Score:	345.50
Percent Similarity:	74.11%
Best local Similarity:	59.82%
Query Match:	51.41%
DB:	3
Length:	105
Matches:	67
Conservative:	16
Mismatches:	20
Indels:	
Gaps:	3

US-09-696-872-24 (1-372) x US-09-124-671-17 (1-105)

QY	10	ATGGGAAGGTACATCTGTTTAA-----GGCTTGCTCGCCCTTGGCGAGTCTGGACGCGT	63
Db	1	MeCglyLysPhePheThrValValAlaAlaAlaLeuLeuLeuGlyAlaVal-----	17
QY	64	GCCAAAAAAGATCCAGCCTGGGTGGAGACTGTGTTCACCTGGGCGCCGAGATGCTT	123
Db	18	ArgAlaIgluglySerSerLeuGlyGlyAspCysCysValAlaLeu-----Val	33
QY	124	CCGGAACCTCAGGAAACCAACCGCGCGTGCAGACGTGGCGGAGATGGCTCGCGCAG	183
Db	34	ThnGlnLeuThrLeuPheAsnGlnLeuValGluLeuAlaGlyAspAspIleAlaArgAspGln	53
QY	184	GTCAGGGAATCACGTTCTCTGAAAAACACGGTGATGAGAGTGACGCGCTCGCGGCGCAG	243
Db	54	ValLysGluMetSerLeuLeuIleArgAsnThrIleMetGlnCysGlnValCysGlyProGln	73
QY	244	CCGCAACCCGAAACCGCAGCCGACGCGCAGCCGACCCGAAACCGCAGCCGAAACCCGAA	303
Db	74	ProGlnProLysProGlnProGlnProGlnProGlnProGlnProLysProGlnProLysProGln	93
QY	304	CCGGAAGGTACCGGATCATCAGAAAAAGTAGTTG	339
Db	94	ProGlnGlyThrIlySerSerGlnLysAspGlnLeu	105

RESULT 10
US-09-124-671-25
; Sequence 25, Application US/09124671A

Alignment Scores:

Pred. No.:	8.54e-23	Length:	96
Score:	304.50	Matches:	6
Percent Similarity:	60.18%	Conservative:	1
Best Local Similarity:	59.29%	Mismatches:	18
Query Match:	45.31%	Indels:	2
DB:	3	Gaps:	2

US-09-696-872-24 (1-372) x US-09-124-671-25 (1-90)

[illegible][illegible]

RESULT 11
US-09-124-671-2
; Sequence 2, Application US/09124671A

Alignment Scores:	
Pred. No.:	1.82e-16
Score:	241.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	35.86%
DB:	3
Gaps:	0
Length:	46
Matches:	46
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-696-872-24 (1-372) X US-09-124-671-2 (1-46)

Qy 100 TCAGACCTTGAGGCCGACAGATGTTGTGGGAACTGACAGAGAAACAACGGCGGCGTGCAGAGAC 159

Db 1 SerAspIeuGClProGImetIeuArgIuIeuGIngtIuThraMalIaIaIeuGInasp 20

Qy 160 GTGCGGGAACTGGCTGCGGACGACAGGTACGGGAGATCACGTTCTCGAAAAACAAGGTGATG 219

Db 21 ValArgAspTrpIeuArgGInGInIaValArgGluIeThrPheIeuIuSaantIuValMet 40

DY	220 GACTGTGACCGCTGGCG	237
Db	41 GLCYSPALACYSGLY	46

RESULT 12
US-09-091-814-46
; Sequence 46, Application US/09091814
; Patent No. 6218513
; GENERAL INFORMATION:
; APPLICANT: Anthony-Cahill, Spencer J

COMPUTER: IBM-compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11725
FILING DATE: filed herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/985,296
FILING DATE: 04-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: GATES, Edward R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: B0801/7005WO
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 889 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: yes
PCT-US93-11725-2

Alignment Scores:
Pred. No.: 9.79e-07 Length: 889
Score: 146.50 Matches: 44
Percent Similarity: 47.33% Conservative: 18
Best Local Similarity: 33.59% Mismatches: 36
Query Match: 21.80% Indels: 33
DB: 5 Gaps: 5

US-09-696-872-24 (1-372) x PCT-US93-11725-2 (1-889)

QY 48 GGCAGTCTGCGAGCCCTGCCAAAAAGATCCAGCTGGTGAGACTGTGTTCAGACCT 107
Db 132 G1yAla11eG1ng1uCysPheMeCg1n-LySeRg1u1aG1ng1nThRg1yAepVa 151
QY 108 GGGCCCGAGATGCTTCGGAGACTGCAGAAACCAACCGCGCTGCAGAGCTGCCGGA 167
Db 151 lSeRArg1n1euR1eG1ng1n1eThRg1MeCAnG1nMeT1euG1yG1uLeuArgAs 171
QY 168 CTGCTGCGGAGACGATCAGGAGATCAGCTTCTGAAAAACACCGTGATGAGTGTA 227
Db 171 pValMeCArg1nG1nVal1ySG1uThMeCPhE1euArgAnThR11eAlaG1uCysG1 191
QY 228 CGCGTGGGGCGCGAGCCGAG---CCGAAACCGAGCCGAGCCGAG----- 273
Db 191 nAlaCysG1y1euG1yProAsPheProLeuProThR1ySValProG1nArgLeuAlaTh 211
QY 274 ---CCGCAGCCGAACCG----- 288
Db 211 rThRThRProR1ySProArgCysAepAlaThSeRcysPheArgG1yValaRgCys11 231
QY 289 -----CAGCCGAAACCGAAACCGAAGT---ACCGATCA----- 321
Db 231 eAsPThRg1uG1yG1PheG1nCysG1yProCysProG1uG1yTyRThRg1yAanG1yVa 251
QY 322 -----TCAGAAAAAGATGAGTTGTAGG 343
Db 251 l11eCysThRAsPValaepG1u-CysArg 260

Search completed: August 26, 2003, 16:26:26
Job time : 17.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_n2p model

Run on: August 26, 2003, 16:24:54 ; Search time 41.5 Seconds

(without alignments)
2365.772 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 672
Sequence: 1 aagcttaccatgggaagta.....aattccatcgcctccgag 372

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 994158

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p -DEV=rlp
-O=/cg21/USPTO.spool/US09696872/runat_26082003_150545_8814/app_query.fasta_1.519
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdd -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HRAPIZ=500 -MINLEN=0
-MALLEN=200000000 -USER=US09696872@cgn_1_1.83@runat_26082003_150545_8814
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA.*

1:	/cg21_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cg21_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cg21_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cg21_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cg21_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cg21_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cg21_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cg21_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cg21_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10:	/cg21_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
11:	/cg21_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
12:	/cg21_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
13:	/cg21_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cg21_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
15:	/cg21_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
16:	/cg21_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
17:	/cg21_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
18:	/cg21_6/ptodata/1/pubpaa/US60_NEW_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	262.5	39.1	242	10	US-09-919-603-5	Sequence 5, Appl1
2	262.5	39.1	300	10	US-09-919-603-7	Sequence 7, Appl1
3	262.5	39.1	757	12	US-09-919-603-3	Sequence 3, Appl1
4	262.5	39.1	757	12	US-10-301-822-41	Sequence 41, Appl1
5	262.5	39.1	757	15	US-10-177-293-72	Sequence 72, Appl1
6	145.5	21.7	961	15	US-10-017-721-4	Sequence 4, Appl1
7	140	20.8	178	15	US-10-154-971-22	Sequence 22, Appl1
8	139	20.7	303	15	US-10-154-971-24	Sequence 24, Appl1
9	126.5	18.8	578	15	US-10-156-761-12544	Sequence 12544, A
10	122	18.2	400	12	US-10-301-822-155	Sequence 155, App
11	118.5	16.9	2456	12	US-10-184-644-57	Sequence 57, Appl1
12	118.5	16.9	2456	15	US-10-184-634-57	Sequence 57, Appl1
13	117	17.4	285	10	US-09-881-752A-228	Sequence 228, Appl1
14	116	17.3	261	15	US-10-280-953-17	Sequence 17, Appl1
15	115.5	16.5	1547	12	US-10-140-472-417	Sequence 417, App
16	115.5	16.5	1547	12	US-10-141-761-417	Sequence 417, App
17	115.5	16.5	1547	12	US-10-123-155-417	Sequence 417, App
18	115.5	16.5	1547	16	US-10-146-731-417	Sequence 417, App
19	115	16.4	2498	12	US-10-140-472-483	Sequence 483, App
20	115	16.4	2498	12	US-10-141-761-483	Sequence 483, App
21	115	16.4	2498	15	US-10-123-155-483	Sequence 483, App
22	115	16.4	2498	16	US-10-146-731-483	Sequence 483, App
23	113.5	16.2	2868	12	US-10-140-472-119	Sequence 119, App
24	113.5	16.2	2868	12	US-10-141-761-119	Sequence 119, App
25	113.5	16.2	2868	15	US-10-123-155-119	Sequence 119, App
26	113.5	16.2	2868	16	US-10-146-731-119	Sequence 119, App
27	113	16.1	1773	15	US-10-184-644-555	Sequence 555, App
28	113	16.1	1773	15	US-10-184-634-555	Sequence 555, App
29	112.5	16.0	3554	12	US-10-140-472-537	Sequence 537, App
30	112.5	16.0	3554	12	US-10-141-761-537	Sequence 537, App
31	112.5	16.0	3554	15	US-10-123-155-537	Sequence 537, App
32	112.5	16.0	3554	16	US-10-146-731-537	Sequence 537, App
33	111.5	15.6	807	14	US-10-138-221-7	Sequence 7, Appl1
34	111.5	15.9	3305	15	US-10-184-644-87	Sequence 87, Appl1
35	111.5	15.9	3305	15	US-10-184-634-87	Sequence 87, Appl1
36	111	15.8	2403	15	US-10-184-644-45	Sequence 45, Appl1
37	111	15.8	2403	15	US-10-184-634-45	Sequence 45, Appl1
38	110.5	15.7	2849	12	US-10-140-472-285	Sequence 285, App
39	110.5	15.7	2849	12	US-10-141-761-285	Sequence 285, App
40	110.5	15.7	2849	15	US-10-123-155-285	Sequence 285, App
41	110.5	15.7	2849	16	US-10-146-731-285	Sequence 285, App
42	110	15.7	2211	14	US-10-096-961-1	Sequence 1, Appl1
43	109.5	15.6	1985	12	US-10-140-472-143	Sequence 143, App
44	109.5	15.6	1985	12	US-10-141-761-143	Sequence 143, App
45	109.5	15.6	1985	15	US-10-123-155-143	Sequence 143, App

ALIGNMENTS

RESULT 1
US-09-919-603-5
Sequence 5, Application US/09919603
Patent No. US20020137679A1
GENERAL INFORMATION:
APPLICANT: Lawler, John W.
TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
TITLE OF INVENTION: Chimeric Proteins
FILE REFERENCE: 1440.1035-007
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/919,603
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/118,053
PRIOR FILING DATE: 1999-02-01
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 242
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chimeric protein
US-09-919-603-5

Alignment Scores:

Pred. No.:	7.86e-16	length:	24
Score:	262.50	Matches:	57
Percent Similarity:	75.61%	Conservative:	5
Best Local Similarity:	69.51%	Mismatches:	9
Query Match:	39.06%	Indels:	11
DB:	10	Gaps:	2

US-09-696-872-24 (1-372) X US-09-919-603-5 (1-242)

Qy	34	TTGCTGCGCCCTTGCGGCACTCTGCAGCGCTGCCAAAAAGATCCAGCCTGGTGGAGAC	93
		: : : : : : : : : : : : : : : : : :	
Db	10	LeuLeuThrLeuAlaIaLeuGlyIaIaSerGlyGlnGlyGlnSerProLeuGly-----	27
Qy	94	TGTTTGTTCAGACTGGGCGCCGACGATGCTTGGGAACTGCAGAGAAACCAACGCGCGCTG	15
		: : : : : : : : : : : : : : : : : :	
Db	28	-----SerAspLeuGlyProGlnMetLeuArgIuLeuGlnGluThrAsnAlaIaLeu	45
Qy	154	CAGACGTCGGGGACTGAGCTGGCGGACGAGCTCAGAGGAGATCAGCTTCTGAAAAACG	213
		: : : : : : : : : : : : : : : : : :	
Db	46	GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleTrpLeuLeuValAsnThr	65
Qy	214	GTGATGAGCTGTGACCGCTGCGGCGCCGACG-----CCGACGCGG	252
		: : : : : : : : : : : : : : : : : :	
Db	66	ValMetGluCysAspAlaCysGlyMetGlnGlnSerValArgThrGlyLeuProSerVal	85
Qy	253	AAACCG	258
Db	86	ArgPro	87

RESULT 2
US-09-91

Sequence 7, Application US/09919603
Patent No. US20020137679A1

```

1  APPLICANT: Lawler, John W.,
2  TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
3  TITLE OF INVENTION: Chimeric Proteins
4  TITLE OF INVENTION: Chimeric Proteins
5  FILE REFERENCE: 1440.1033-007
6  CURRENT APPLICATION NUMBER: US/09/919,603
7  PRIOR FILING DATE: 2001-07-30
8  PRIOR APPLICATION NUMBER: PCT/US00/02482
9  PRIOR FILING DATE: 2000-02-01
10 PRIOR APPLICATION NUMBER: 60/118,053
11 PRIOR FILING DATE: 1999-02-01
12 NUMBER OF SEQ. ID NOS: 21
13 SOFTWARE: FastSeq for Windows Version 4.0
14 SEQ. ID NO 7
15 LENGTH: 300
16 TYPE: PRT
17 ORGANISM: Artificial Sequence
18 FEATURE:
19 OTHER INFORMATION: chimeric protein
20 US-09-919-603-7

```

Alignment Scores:

Pred. No.:	8.04e-16	Length:	300
Score:	262.50	Matches:	57
Percent Similarity:	75.61%	Conservative:	5
Best Local Similarity:	69.51%	Mismatches:	9
Query Match:	39.06%	Indels:	1
DB:	10	Gaps:	2

US-09-696-872-24 (1-372) x US-09-919-603-7 (1-300)

QY 3 TTGCTGCGCCCTTGGCGGCATCTCTGACGGCTGCCAAAAAGATCCAGCTGGGTGGAGAC 93
 Db 10 LeuLeuThrLeuAlaIaLeuGtYAlaSerGtYgInGtYInSerProLeuGtY----- 27
 QY 94 TGTTCCTTCAGACTCTGGGCGCCGACGATGCTTGGGAACTGCAGAGAAACCAACGCGCGCTG 153
 Db 28 -----SerAspLeuGtYProGlnInnetLeuAlaGtYtulebuGtYInGtYThrAspAlaIaLeu 45
 .

QY 154 CAGGACGTGCGGGAACGCGCTGCGGAGGACAGGTCACGGAGATCAGCTTCTGAAAAACAG 213
 Db 46 GlnAspValAlaGAspTyrLeuArgGlnGlnValAlaGluuIleThrPheLeuIleAsnThr 65
 QY 214 GTGATGAGTGTGACCGCTGCGGCGCCGAG-----CGGACCG 255
 Db 66 ValMetGluCysAspAlaCysGlnIleGlnGlnValAlaArgThrGlyLeuProSerVal 85

RESULT 3
US-09-919-603-3

Sequence 3, Application US/09919603
Patent No. US20020137679A1
CURRENT INFORMATION

```

? APPLICANT: Lawler, John W.
? TITLE OF INVENTION: COMP/TSB-1, COMP/TSB-2 and Other TSB
? TITLE OF INVENTION: Chimeric Proteins
? FILE REFERENCE: 1440.1035-007
? CURRENT APPLICATION NUMBER: US/09/919, 603
? PRIOR FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: PCT/US00/02482
? PRIOR FILING DATE: 2000-02-01
? PRIOR APPLICATION NUMBER: 60/118, 053
? PRIOR FILING DATE: 1999-02-01
? NUMBER OF SEQ ID NOS: 21
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 3
? LENGTH: 757
? TYPE: PRT
? ORGANISM: Homo sapiens
? OS-09-919-603-3

```

Alignment Scores:

Pred. No.:	8.85e-16	Length:	75
Score:	263.50	Matches:	57
Percent Similarity:	75.61%	Conservative:	5
Best Local Similarity:	69.51%	Mismatches:	9
Query Match:	39.06%	Indels:	11
DB:	10	Gaps:	2

US-09-696-872-24 (1-372) x US-09-919-603-3 (1-757)

QY	34	TTGCTGCGCCCTTGCGGCAAGTGTGCAGCGCCGTGCCAAAAAGAGATCCAGCCTGGGTGGAGAC	93
	10	LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyGlnGlyGlnSerProLeuGlyY-----	27
Db	94	TGTTGTTGAGACTTGCGGCCGACAGATGTTGGGGAATGCAGAAACCAACGCGCGCTG	153
QY	28	-----SerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGlnThrAsnAlaAlaLeu	45
Db	154	CAGAGCTGCGGGACTGTGGCTGCGGACAGAGTCAAGGAGATCACGTTCTGAAAAACAG	213
QY			
Db	46	GlnAspValArgAspPrlPheArgGlnGlnValAlaArgGlnIleThrPheLeuLysAsnThr	65
QY	214	GTGATGAGTGTGAAGCGCTGCGGCGCCGAG-----CCGACGCCG	252
Db	66	ValMetGlnCysAspAlaCysGlnMetClnGlnSerValArgThrGlyLeuProSerVal	85
QY	253	AAACCG	258
Db	86	ArgPro	87

RESULT 4

US-10-301-822-41
; Sequence 42, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc
; APPLICANT: Berger, Allison

```

; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MPM01-0292RRM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-10-301-822-41

Alignment Scores:
Pred. No.: 8,85e-16 Length: 757
Score: 262.50 Matches: 57
Percent Similarity: 75.61% Conservative: 5
Best Local Similarity: 69.51% Mismatches: 9
Query Match: 39.06% Indels: 11
DB: 12 Gaps: 2

US-09-696-872-24 (1-372) x US-10-301-822-41 (1-757)
QY 34 TTGTCGCCCTTGCGGAGCTGTCAGCGCTGCCAAGATCCAGCTGGGTGAGAC 93
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
10 LeuettThrluAlaAlaLeuGlYAlaSerGlYngInglYnSerProLeuGlY----- 27
QY 94 TGTGTTCAAGACTGGGCGCGGACAGATGCTTGGGAACTGCAGAAACCAACGCGCGCTG 153
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28 -----SerAspLeuGlYProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 45
QY 154 CAGACGTCGGGAGCTGGCTGGCGAGAGAGTCAGGAGATCAGCTTCTGAAAAACAGC 213
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuThrPheLeuLysAsnThr 65
QY 214 GTGATGAGTGTGACGCGCTGGCGGCGCGAG-----CCGACGCGC 252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 ValMetGluCyAspAlaCySglYMetGlnGlnSerValArgThrGlyLeuProSerVal 85
QY 253 AAACCG 258
Db |||||
86 ArgPro 87

RESULT 5
US-10-177-293-72
; Sequence 72, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lilije, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongxiao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.

```

```

; APPLICANT: Baet Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzstai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sanin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-177-293-72

Alignment Scores:
Pred. No.: 8,85e-16 Length: 757
Score: 262.50 Matches: 57
Percent Similarity: 75.61% Conservative: 5
Best Local Similarity: 69.51% Mismatches: 9
Query Match: 39.06% Indels: 11
DB: 15 Gaps: 2

US-09-696-872-24 (1-372) x US-10-177-293-72 (1-757)
QY 34 TTGTCGCCCTTGCGGAGCTGTCAGCGCTGCCAAGATCCAGCTGGGTGAGAC 93
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
10 LeuettThrluAlaAlaLeuGlYAlaSerGlYngInglYnSerProLeuGlY----- 27
QY 94 TGTGTTCAAGACTGGGCGCGGACAGATGCTTGGGAACTGCAGAAACCAACGCGCGCTG 153
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28 -----SerAspLeuGlYProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 45
QY 154 CAGACGTCGGGAGCTGGCTGGCGGAGAGTCAGGAGATCAGCTTCTGAAAAACAGC 213
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
46 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuThrPheLeuLysAsnThr 65
QY 214 GTGATGAGTGTGACGCGCTGGCGGCGCGAG-----CCGACGCGC 252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 ValMetGluCyAspAlaCySglYMetGlnGlnSerValArgThrGlyLeuProSerVal 85
QY 253 AAACCG 258
Db |||||
86 ArgPro 87

RESULT 6
US-10-017-721-4
; Sequence 4, Application US/10017721
; Publication No. US20030096248A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Daley, George
; APPLICANT: Bolik, Stacey
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMI-003
; CURRENT APPLICATION NUMBER: US/10/017,721
; CURRENT FILING DATE: 2001-12-14

```

;; PRIOR APPLICATION NUMBER: US 60/317,033
;; PRIOR FILING DATE: 2001-09-04
;; PRIOR APPLICATION NUMBER: US 60/330,248
;; PRIOR FILING DATE: 2001-10-17
;; NUMBER OF SEQ ID NOS: 13
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 4
;; LENGTH: 961
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-017-721-4

Alignment Scores:
Pred. No.: 4,376-05 Length: 961
Score: 145.50 Matches: 30
Percent Similarity: 57.14% Conservative: 10
Best Local Similarity: 42.86% Mismatches: 25
Query Match: 21.65% Indels: 5
DB: 15 Gaps: 2

US-09-696-872-24 (1-372) x US-10-017-721-4 (1-961)

QY 103 GACCTGGGCGCCGACATGCTTGGGAGACTGCAGAAACCAAGCGCGCTGCAGACGCG 162
DB 218 AaPheaaMaArgGlnPheLeuGlyGlnMetThrGlnLeuAaGlnLeuLeuGlyGlnVal 237
QY 163 CGGAGCTGCTGCGGAGCAGCAGGTGAGATCACGTTCTGAAAAACAGCGTGTAGAG 222
DB 228 LysaAaPheLeuAaArgGlnGlnValLysGlnThrSerPheLeuAaArgAaThrLLeaGlu 257
QY 223 TGTGACGCTGCGGCGCG-----CAGCCGACAGCCGAAACCCGACGCGACGCG 270
DB 258 CysGlnAaCyGgLyPheLeuLysPheGlnSerProThrPheValValaLaPro 277
QY 271 CAGCCGACGCGAAACCGACGCGAAACCG 300
DB 278 AlaPro---ProAlaProProThrArgPro 286

RESULT 7
US-10-154-971-22
; Sequence 22, Application US/10154971
; Publication No. US20030088074A1
; GENERAL INFORMATION:
; APPLICANT: Hamers, Raymond
; Myldermans, Serge
; TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -
; USE FOR THERAPEUTIC OR VETERINARY PURPOSES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESSES:
; ADDRESS: SPENCER & FRANK
; STREET: 1100 New York Avenue, N.W., Suite 300 East
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/154,971
; FILING DATE: 28-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/945,244
; FILING DATE: <Unknown>
; APPLICATION NUMBER: PCT/EP96/01725
; FILING DATE: 25-APR-1996
; APPLICATION NUMBER: EP 95400932.0
; FILING DATE: 25-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gollin, Michael A.

;; REGISTRATION NUMBER: 31,957
;; REFERENCE/DOCKET NUMBER: GUPLA 0003
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-414-4000
;; TELEFAX: 202-414-4040
;; INFORMATION FOR SEQ ID NO: 22:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 178 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-10-154-971-22

Alignment Scores:
Pred. No.: 0,000117 Length: 178
Score: 140.00 Matches: 24
Percent Similarity: 88.89% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 3
Query Match: 20.83% Indels: 0
DB: 15 Gaps: 0

US-09-696-872-24 (1-372) x US-10-154-971-22 (1-178)

QY 228 CCGCAGCCGACCGAAACCGCAGCCGACCGCAGCCGAAACCGCAGCCGAAA 297
DB 138 ProGlnProGlnProLyPProGlnProGlnProGlnProGlnProLyPProGlnProLyS 157
QY 298 CCGGAACCGGAAGGTACCGGA 318
DB 158 ProGlnProMetGluCyGgLy 164

RESULT 8
US-10-154-971-24
; Sequence 24, Application US/10154971
; Publication No. US20030088074A1
; GENERAL INFORMATION:
; APPLICANT: Hamers, Raymond
; Myldermans, Serge
; TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -
; USE FOR THERAPEUTIC OR VETERINARY PURPOSES
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESSES:
; ADDRESS: SPENCER & FRANK
; STREET: 1100 New York Avenue, N.W., Suite 300 East
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/154,971
; FILING DATE: 28-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/945,244
; FILING DATE: <Unknown>
; APPLICATION NUMBER: PCT/EP96/01725
; FILING DATE: 25-APR-1996
; APPLICATION NUMBER: EP 95400932.0
; FILING DATE: 25-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gollin, Michael A.
; REGISTRATION NUMBER: 31,957
; REFERENCE/DOCKET NUMBER: GUPLA 0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-414-4000
; TELEFAX: 202-414-4040
; INFORMATION FOR SEQ ID NO: 24:

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 26, 2003, 16:18:39 ; Search time 29.5 Seconds
(without alignments)
2425.407 Million cell updates/sec

Title: US-09-696-872-24
Perfect score: 672
Sequence: 1 aagttaccatgggaagta.....aattccatcgcctcgcag 372

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=rlp
-Q=/cgn2/1/USFTO.spool/US09696872/runat_26082003_150543_8735/app.query.fasta_1.519
-DB=PIR_76 -QFMT=faeaaan -SUFFIX=trp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODB=LOCAL
-CURFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09696872 @CNC 1.1 62 @runat_26082003_150543_8735 -NCP=6 -ICU=3
-NO MMAR -LARGQUTRY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	234	34.8	755	2 A44315	cartilage oligomer
2	146.5	21.8	955	2 A45441	chondrospindin 4 -
3	146	21.7	135	2 S33598	Ig gamma-2 chain -
4	145.5	21.7	961	1 TSHUP4	thrombospondin 4 p
5	133	19.8	332	1 OZ2QMB	circumsporozoite p
6	133	19.8	348	1 OZ2QMB	circumsporozoite p
7	132	19.6	3164	1 WMBE66	ULB6 protein - hum
8	124	18.5	1494	2 T14355	protein-tyrosine-p
9	122	18.2	274	2 A53335	myelin regulatory
10	122	18.2	400	2 S58222	PQ-rich protein -
11	119.5	17.8	129	2 A26036	procyelic acidic r
12	119	17.7	165	2 T24470	hypothetical prote
13	119	17.0	260	2 T17638	glycine tyrosine-r
14	118.5	17.6	378	2 S14959	proline-rich prote

15	118	17.6	422	2 C91067	hypothetical prote
16	117	17.4	285	2 B64687	siderophore-medi
17	117	17.4	309	2 T17557	procyelin homol
18	117	17.4	497	2 F83634	hypothetical prote
19	116.5	17.3	115	1 PAUTPC	procyelic acidic r
20	116.5	17.3	143	2 A26918	procyelic acidic r
21	116.5	17.3	145	2 A44418	siderophore-medi
22	115	17.1	280	2 F71829	com protein limpo
23	114	17.0	266	2 T44781	hypothetical prote
24	113.5	16.9	127	2 S09372	glutamate/proline
25	113.5	16.9	139	2 S70010	glutamate/proline
26	113.5	16.9	221	2 S70009	glutamate/proline
27	113	16.8	407	2 S82478	hypothetical prote
28	112.5	16.7	839	1 TQZMCA	probable transpos
29	112.5	16.7	949	2 T08658	hypothetical prote
30	112.5	16.7	1528	2 D85912	hypothetical prote
31	112.5	16.7	1569	2 A65044	hypothetical prote
32	112.5	16.7	1571	2 C91068	hypothetical prote
33	112	16.7	141	2 T09546	extensin like prot
34	112	16.7	283	2 T16348	hypothetical prote
35	112	16.7	907	2 T15792	hypothetical prote
36	111.5	16.6	138	2 B27863	Ac hypothetical pr
37	111.5	16.6	807	2 T02916	hypothetical prote
38	111	16.5	427	2 T03955	probable transpos
39	111	16.5	1749	2 S75138	hypothetical prote
40	110.5	16.4	149	2 B96651	protein T3P18.7 (i
41	110.5	16.4	179	2 T14145	hypothetical prote
42	110	16.4	792	2 F90566	conserved hypotet
43	109.5	16.3	221	2 B42701	PR264 protein - ch
44	109	16.2	197	2 T18918	hypothetical prote
45	109	16.2	312	2 T26085	hypothetical prote

ALIGNMENTS

RESULT 1
A44315
N:Alternate names: thrombospondin homol COMP
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Apr-1993 #sequence_revision 03-May-1994 #text_change 02-Aug-2002
C:Accession: A44315
R:Oldberg, A.; Antonsen, P.; Lindblom, K.; Heinegard, D.
J. Biol. Chem. 267, 22346-22350, 1992
A:Title: COMP (cartilage oligomeric matrix protein) is structurally related to the throm
A:Reference number: A44315; MUID:93054522; PMID:1429587
A:Accession: A44315
A:Molecule type: mRNA
A:Residues: 1-755 <OLD>
A:Cross-references: GB:X72914; NID:G297438; PIDN:CA51419.1; PID:G297439
A:Experimental source: tracheal chondrocyte
A>Note: Sequence extracted from NCBI Backbone (NCBI:117022)
C:Superfamily: thrombospondin 3; EGF homology
C:Keywords: pentamer
F:1-19/Domain: signal sequence #status predicted <SIG>
F:182-219/Domain: EGF homology <EGF>

Alignment Scores:
Pred. No.: 6.48e-13 Length: 755
Score: 234.00 Matches: 51
Percent Similarity: 72.84% Conservative: 8
Best Local Similarity: 62.96% Mismatches: 18
Query Match: 34.82% Indels: 4
DB: 2 Gaps: 1

US-09-696-872-24 (1-372) x A44315 (1-755)
QY 34 TTGCTGGCCCTTGGGAGCTGCGAGCGCTGCCAAAAGATCCAGCTGGTGAGAC 93
DB 9 Valledalalevalalevalargalathrgllyngllyglnlleproleuglyglly--- 27
QY 94 TGTGTTTCAGACCTGGGCGCCAGATGCTTGGGGAATGCGAGAAACCAAGCGGCGCTG 153

```

Db -----AspleuAlaProGlnMetLeuArgGluLeuGlnGluThrAsnAlaLeu 44
QY 154 CAGACGTGCGGAGCTGCTGCGGAGCAGTCCAGGAGATCAGTTCCTGAAAAACAGC 213
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 45 GlnApyValArgGluLeuLeuArgHisArgVallylGluIleThrPheLeuYsaThr 64
QY 214 GTGATGAGTGTGACGCGTGGCGGCGCGACCGGAAACCGCAGCCGACCGCAG 273
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 65 ValMetGluCyAspAlaCyAsGlyMetGlnProAlaArgThrProGlyLeuSerValArg 84
QY 274 CCG 276
    |||
Db 85 Pro 85

RESULT 2
A45441
Thrombospondin 4 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000
C:Accession: A45441
R:Lawler, J.; Duquette, M.; Whitaker, C.A.; Adams, J.C.; McHenry, K.; Desimone, D.W.
J. Cell Biol. 120, 1059-1067, 1993
A:Title: Identification and characterization of thrombospondin-4, a new member of the th
A:Reference number: A45441; M01D:93163109; PMID:8432726
A:Accession: A45441
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-955 <LAW>
A:Cross-references: GB:Z19091; NID:9288777; PIDN:CA479518.1; PID:9288778
A>Note: sequence extracted from NCBI backbone (NCBIN:124858, NCBI:124860)
F:285-319/Domain: EGF homology <EGF>

Alignment Scores:
Pred. No.: 3 3e-05 Length: 955
Score: 146.50 Matches: 44
Percent Similarity: 47.33% Conservative: 18
Best Local Similarity: 33.59% Mismatches: 36
Query Match: 21.80% Indels: 33
DB: 2 Gaps: 5

US-09-696-872-24 (1-372) x A45441 (1-955)
QY 48 GGCAGCTGTGACGCGTCCAAAAAGATCCAGCTGTGAGAGTGTGTTTCAGACCT 107
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 198 GylAlaIleGlnGluCyPheMetGln-LysSerGluAlaGlyGlnGlnThrGlyAspVa 217
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 108 GGGCCCGCAGATGCTTCGGGAACTGCAGGAAACCAACGGCGCTGCAGGAGCTGGCGGA 167
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 217 IserArgGlnLeuIleGlyGlnIleThrGlnMetAsnGlnMetLeuGlyGluLeuArgAs 237
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 168 CTGGCTGGCGGAGAGTCCAGGAGATCAGTTCCTGAAAAACAGCGATGTGAGTGA 227
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 237 PValMetArgGlnGlnVallylGluIleThrMetPheLeuArgAsnThrIleAlaGluCySeI 257
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 228 CGCGTGGCGGCGCAGCCGACG---CCGAAACCGCAGCCGACCGCAG--- 273
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 257 nAlaCyAsGlyLeuGlyProAspPheProLeuProThrLysValProGlnArgLeuAlaTh 277
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 274 ---CCGACGCGGAACCG--- 288
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 277 rThrThrProProLysProArgCyAspAlaThrSerCyPheArgGlyValArgCyS11 297
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 289 ---CAGCCGAAACCGGAACCGGAAGGT---ACCGATCA----- 321
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 297 eaSpThrGluGlyGlyPheGlnCyAsGlyProCySProGluGlyTyTrnGlyAsnGlyVa 317
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 322 ---TCAGAAAAAGATGACTGTAGG 343
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 317 lIleCyThrAspValAspGlu-CysArg 326
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 3
S33598*

```

```

Ig gamma-2 chain - Arabian camel (fragment)
C:Species: Camelus dromedarius (Arabian camel)
C:Date: 19-Mar-1997 #sequence_revision 10-Oct-1997 #text_change 11-Jan-2000
C:Accession: S33598
R:Hamers-Casterman, C.; Alartouch, T.; Muyldermans, S.; Robinson, G.; Hamers, C.; Songa,
Nature 363, 446-448, 1993
A:Title: Naturally occurring antibodies devoid of light chains.
A:Reference number: S33598; MUID:93275410; PMID:8502296
A:Accession: S33598
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-135 <HAM>
A:Experimental source: spleen
C:Superfamily: immunoglobulin V region, immunoglobulin homology
C:Keywords: disulfide bond; immunoglobulin
F:14-62/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 5 07e-05 Length: 135
Score: 146.00 Matches: 25
Percent Similarity: 96.15% Conservative: 0
Best Local Similarity: 96.15% Mismatches: 1
Query Match: 21.73% Indels: 0
DB: 2 Gaps: 0

US-09-696-872-24 (1-372) x S33598 (1-135)
QY 238 CCGGACCGGACCGGAAACCGGACCGGACCGGACCGGAAACCGGACCGGAA 297
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 91 ProGlnProGlnProLysProGlnProGlnProGlnProGlnProGlnProLys 110
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 298 CCGGAACCGGAAGGTACC 315
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 111 ProGlnProGlnCyThr 116
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 4
TSHUP4
Thrombospondin 4 precursor - human
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 11-Aug-1995 #text_change 13-Aug-1999
C:Accession: A55710; S36069
R:Lawler, J.; McHenry, K.; Duquette, M.; Derick, L.
J. Biol. Chem. 270, 2809-2814, 1995
A:Title: Characterization of human thrombospondin-4.
A:Reference number: A55710; MUID:95155352; PMID:7852353
A:Accession: A55710
A:Molecule type: mRNA
A:Residues: 1-961 <LAW>
A:Cross-references: EMBL:Z19585; NID:9311625; PIDN:CA479635.1; PID:9311626
A>Note: authors translated the codon GTG for residue 616 as Ser
C:Genetics:
A:Gene: GDB:THB84
A:Cross-references: GDB:463011; OMIM:600715
A:Map position: 1q21-1q23
C:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregation
C:Superfamily: thrombospondin 3; EGF homology
C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; cell adhe
F:121-261/Product: thrombospondin 4 #status predicted <SIG>
F:290-324/Domain: EGF homology <EGF>
F:330-362/Domain: EGF homology <EGF>
F:562-564/Region: cell attachment (R-G-D) motif
F:303/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F:343/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:612/941/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 4 04e-05 Length: 961
Score: 145.50 Matches: 30
Percent Similarity: 57.14% Conservative: 10
Best Local Similarity: 42.86% Mismatches: 25

```


A26036
 procytic acidic repetitive protein A-beta - Trypanosoma brucei
 N:Alternate names: Insect-stage-specific protein
 C:Species: Trypanosoma brucei
 C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 13-Aug-1999
 C:Accession: A26036; S21540
 R:Roditi, I.; Carrington, M.; Turner, M.
 Nature 325, 272-274, 1987
 A:Title: Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*
 A:Reference number: A26036; MUID:87115776; PMID:3808022
 A:Accession: A26036
 A:Molecule type: mRNA
 A:Residues: 1-129 <R0D>
 A:Cross-references: GB:Q4814; NID:q10510; PIDN:CAA28503.1; PID:q10511
 R:Vijayasarathy, S.; Ernest, I.; Itzhaki, J.; Sherman, D.; Mowatt, M.R.; Michals, P.A.M. submitted to the EMBL Data Library, April 1990
 A:Description: The genes encoding fructose biphosphate aldolase in *Trypanosoma brucei* A:Reference number: S21538
 A:Accession: S21540
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-129 <VLD>
 A:Cross-references: EMBL:X52584; NID:q10480; PIDN:CAA36815.1; PID:q10482
 C:Superfamily: procytic acidic repetitive protein
 F:63-106/Region: 2-residue repeats (E-P)

Alignment Scores:
 Pred. No.: 0.0112 Length: 129
 Score: 119.50 Matches: 25
 Percent Similarity: 58.82% Conservative: 15
 Best Local Similarity: 36.76% Mismatch: 16
 Query Match: 17.78% Indels: 12
 DB: 2 Gaps: 3

US-09-696-872-24 (1-372) x A26036 (1-129)

QY 108 GCGCCCGCAGATGCTTCGCGAATCGCAGAAACCAACCGCGCTCGCAGCATGCGCGGA 167
 Db 30 GYLPProgiuAap-----LysGlyLeuThrLySGlyGlyLys----- 43
 QY 168 CTGGCTGGCGAGCAGCGTCAAGGAGATCAGCTCTCTGAAAAACCGTGAATGAGTGTGA 227
 Db 44 -----GlyGlyLysGlyThrLySValSerAspAspAspThrAsnGlyThr-As 59
 QY 228 CGCGTGGCGGCGCGAGCCGAGCCGAAACCGCAGCGCGCAGCGCGCAGCGCGAACC 287
 Db 59 pPro---AspProgiuPProgiuPProgiuPProgiuPProgiuPProgiuPProgiuP 78
 QY 288 CGAGCCGAAACCGGAAACCGGAA 309
 Db 78 OGluPProgiuPProgiuPProgiu 85

RESULT 12
 T24470
 hypotheoretical protein T04F8.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T24470
 R:Lennard, N.
 submitted to the EMBL Data Library, November 1995
 A:Reference number: Z19895
 A:Accession: T24470
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-165 <NLL>
 A:Cross-references: EMBL:Z66565; PIDN:CAA91483.1; GSPDB:GN00028; CESP:T04F8.8
 A:Experimental source: clone T04F8
 C:Genetics:
 A:Gene: CESP:T04F8.8
 A:Map position: X
 A:Introns: 18/2; 63/1

Alignment Scores:

```

Pred. No.: 0 0119 Length: 165
Score: 119.00 Matches: 28
Percent Similarity: 44.44% Conservative: 8
Best Local Similarity: 34.57% Mismatches: 28
Query Match: 17.71% Indels: 17
Db: 2 Gaps: 2

US-09-696-872-24 (1-372) x T24470 (1-165)

QY 99 TTCGAGCTGGGCGCGCAGATGCTTCGGGAACTCGAGGAAACCAACGCGGCGCTGCAGCA 158
   ::|||
Db 66 TyAaNaAmnGlyProProProPhaeapGlyGlyTyTzAaArGProTyrgLyGlyGly 85
QY 159 CGTGGCGGACTGGCTGGCGGACGAGT-----CAG 168
   |||
Db 86 GlyGlyGlyTyTzGlyArGProGlnGlyTyTzAlProProProProProProThGln 105
QY 189 GGAGATCAAGTTCCTGGAAAAACACGGGTATGAGAGTGCAGCGCGGCGCGGACCGCA 248
   ::|||
Db 106 ProGlnProGlnProGlnPro-----ArGProGlu-ProGlnProGln 119
QY 249 GCCGAACCGCAGCCGACCGCAGCCGACCGCGAAACCGCAGACCGAACCGGA 308
   :|||||:|||||:|||||:|||||:|||||:|||||
Db 119 uProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProHsProGln 139

QY 309 A 309
   |
Db 139 u 139

RESULT 13
T17638
glycine tyrosine-rich protein a147L - Chlorella virus PBCV-1
C:/Species: Chlorella virus PBCV-1
C:/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:/Accession: T17638
R:/Graves, M.V., Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:/Reference number: Z18806
A:/Accession: T17638
A:/Status: preliminary; translated from GB/EMBL/DBJ
A:/Molecule type: DNA
A:/Residues: 1260 <GRA>
A:/Cross-references: EMBL:U42580, NID:g4028896, PIDN:AAC96515.1
C:/Genetics:
A:/Experimental source: specific host Chlorella strain NC64A
A:/Note: a147L

Alignment Scores:
Pred. No.: 0.011 Length: 260
Score: 119.00 Matches: 20
Percent Similarity: 80.00% Conservative: 0
Best Local Similarity: 80.00% Mismatches: 3
Query Match: 16.95% Indels: 2
Db: 2 Gaps: 1

US-09-696-872-24 (1-372) x T17638 (1-260)

QY 305 GGTTCGGGTTTCGGCGCGCGGTTTCGAGTGC-----GGCTGCGGCTGGCGGCTGGTTTC 252
   |||
Db 67 GlyCyGgGlyPheGlyYCyGgGlyPheGlyYCyGgGlyPheGlyYCyGgGlyYCyGgGlyPhe 86
QY 251 GGCTGCGGCTGGCGCG 237
   |||
Db 87 GlyCyGgGlyPheGly 91

RESULT 14
S14959
proline-rich protein - wheat
C:/Species: Triticum aestivum (common wheat)
C:/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Nov-1999
C:/Accession: S14959
C:/Graves, C.A.; Lloyd, J.C.; Chao, S.; John, U.P.; Murphy, G.-J.P.
Plant Mol. Biol. 16, 663-670, 1991

```

A:Title: A novel proline-rich protein from wheat
A:Reference number: S14959; NCID:91329659; PMID:1714320
A:Accession: S14959
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-378 <RAI>
A:Cross-references: EMBL:X52472; NID:g21841; PIDD:CAA36712.1; PID:g218442
C:Superfamily: hydroxyproline-rich glycoprotein

Search completed: August 26, 2003, 16:25:51
Job time : 33.5 secs

Alignment Scores:	
Pred. No.:	0.014
Score:	118.50
Percent Similarity:	59.70%
Best Local Similarity:	43.28%
Query Match:	17.63%
DB:	2
Length:	378
Matches:	29
Conservative:	11
Mismatches:	23
Indels:	4
Gaps:	2

US-09-696-872-24 (1-372) X SI4959 (1-378)

QY	108	GGGCGCCGCAAGATGCTTCGGGAACTGCAGAGAAACAAGCGGCGTGCAGAGACGCGCGGA	167
Db	13	GlyLeuValAlaIaIaSerGlyPhe-----SerGlnAlaAlaAlaAlaGlyArg--Gly	29
QY	168	CTGGCTGCGGACGACAGTGCAGGAGATCACGTTCCCTGAAAAACAACGCGTGAATGAGAGTGTGA	227
Db	30	LeuAlaGluLysLeuProGluProGluProLysProThrProCysProGluProLysPro	49
		:::	:::
		:::	:::
		:::	:::
QY	228	CGCGTGCGGGCGCGACGCCGCGAACCAGCAGCCGACGCGACGCGGCGAGCCGAAACC	287
		:::	:::
		:::	:::
		:::	:::
Db	50	GlnProLys-ProGluProMetProLysProGluProMetProLysProGluProLysPro	69
		:::	:::
		:::	:::
		:::	:::
QY	288	GCAGCGGAAACCGGAAACCG	306
Db	69	oleuProLysProGluPro	75

RESULT 15

hypochlorite protein ECg3507 [imported] - *Escherichia coli* (strain O157:H7, substrain RHD)
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 A:Accession: C91067
 R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 Nucleic Acids Res. 9, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C91067
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-422 <HAY>
 A:Cross-references: GB:BA000007; PIDD:BA836930.1; PIDD:g13362978; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 050952
 C:Genetics:
 C:Gene: ECg3507

Alignment Scores:	
Pred. No.:	0.0124
Score:	118.00
Percent Similarity:	87.50%
Best Local Similarity:	53.12%
Query Match:	17.56%
DB:	2
Length:	422
Matches:	17
Conservative:	11
Mismatches:	4
Indels:	1
Gaps:	0

US-09-696-872-24 (1-372) X C91067 (1-422)

QY	238	CCGCACCCGAGCCGAAA	CCGCAG	CGCAG	CGCAG	CGCGAACCGAAACGCACCGAA	297
Db	367	ProGluProGluProGluProGluProGluProGluProGluProGlu	ProGluProGluProGluProGluProGluProGluProGluProGlu	ProGluProGluProGluProGluProGluProGluProGluProGlu	ProGluProGluProGluProGluProGluProGluProGluProGlu	ProGluProGluProGluProGluProGluProGluProGluProGlu	366
QY	298	CCGGAACCGGAAGTTC	CGGATCATCAGAAAAAT	333			
Db	387	ProGluProGluProLeuIeArgSerLeuNyGlu	ProGluProGluProLeuIeArgSerLeuNyGlu	398			

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: August 26, 2003, 16:07:14 ; Search time 17.5 Seconds
(without alignments)
1999.307 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 672
Sequence: 1 aagcttaccatgggaagta.....attccatcatgcattcgag 372

Scoring table:

	BLOSUM62
Xgapop 10.0 , Xgapext 0.5	
Xgapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_nzp.model -DEV=xlp
-O=/cgm2.1/USPTO_epool/US09696872/runat_26082003.150542.8719/epg_query.fasta.1.519
-DB=swissprot 41 -QFMT=fastan -SUFFIX=rsb -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blissum62 -TRANS=human40.cdi -LIST=45
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09696872 @CGN 1.1.30 @runat 26082003.150542.8719 -NCP=6 -ICU=3
-NO MAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=110 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	262.5	39.1	757	1 COMP_HUMAN	P49747 homo sapien
2	234	34.8	755	1 COMP_RAT	P35444 rattus norv
3	146.5	21.8	955	1 TSP4_XENLA	Q06441 xenopus lae
4	145.5	21.7	961	1 TSP4_HUMAN	P35443 homo sapien
5	144.5	21.5	980	1 TSP4_RAT	P49744 rattus norv
6	133	19.8	339	1 CSP_PLAAB	P06915 plasmodium
7	133	19.8	347	1 CSP_PLAAB	P23093 plasmodium
8	132	19.6	3164	1 TEGU_HSV1	P10220 herpes simp
9	124	18.5	220	1 NOJ3_MOUSE	Q94100 mus musculi
10	122	18.2	283	1 SPFA_MOUSE	Q06701 mus musculi
11	119.5	17.8	129	1 PARB_TRYBB	P09791 trypanosoma
12	117	17.4	285	1 TONB_HELPY	O25899 helicobacte
13	116.5	17.3	115	1 PARX_TRYBB	P14043 trypanosoma
14	116.5	17.3	143	1 PAR1_TRYBB	P08469 trypanosoma
15	116.5	17.3	145	1 PARC_TRYBB	Q06084 trypanosoma
16	116	16.5	194	1 KRUB_HUMAN	O75690 homo sapien
17	115	17.1	280	1 TONB_HELPY	Q94100 mus musculi
18	113.5	16.9	221	1 NOJ3_RAT	Q62881 rattus norv

19	112.5	16.7	839	1 TRA9_MAIZE	P03010 zea mays (m
20	112.5	16.7	1569	1 YPUA_ECOLI	P52143 escherichia
21	111	16.5	806	1 TRAI_MAIZE	P08770 zea mays (m
22	109.5	16.3	221	1 SFR2_CHICK	P30352 gallus gall
23	108	16.1	722	1 Z219_HUMAN	Q9P294 homo sapien
24	107	15.9	503	1 VE2_HPV21	P50767 human papil
25	107	15.9	613	1 PKPA_PHYBL	Q01577 phycomyces
26	106	15.8	955	1 T150_HUMAN	O9Y2W1 homo sapien
27	105.5	15.0	131	1 CHHB_BOMMO	P05688 bombyx mori
28	105	15.6	238	1 SFR7_HUMAN	Q16629 homo sapien
29	105	15.6	243	1 TONB_KLEBN	P45610 klebsiella
30	104.5	15.6	221	1 SFR2_HUMAN	Q01130 homo sapien
31	104.5	15.6	221	1 SFR2_MOUSE	O62093 mus musculi
32	104.5	15.6	482	1 UZR2_HUMAN	O15696 homo sapien
33	104.5	15.6	1205	1 S122_MOUSE	P55012 mus musculi
34	103.5	15.4	375	1 SRS5_DROME	P26686 drosophila
35	103.5	15.4	483	1 VE2_HPV14	P36783 human papil
36	103.5	15.4	493	1 VE2_HPV19	P36786 human papil
37	103.5	15.4	1493	1 M3K1_RAT	O62925 rattus norv
38	103	15.3	269	1 SFR5_RAT	O09167 rattus norv
39	102.5	15.3	494	1 SFR4_HUMAN	O08170 homo sapien
40	102.5	15.3	497	1 VE2_HPV20	P50766 human papil
41	102	15.2	449	1 APG_BRANA	P40603 brassica na
42	101	15.0	270	1 SFR5_MOUSE	O35326 mus musculi
43	101	15.0	342	1 TONB_PSEAE	O51368 pseudomonas
44	101	15.0	450	1 V50K_BYDVP	P09516 barley yell
45	101	15.0	565	1 MOT8_MOUSE	O70324 mus musculi

ALIGNMENTS

RESULT 1

ID	COMP_HUMAN	STANDARD	PRT	757 AA.
AC	P49747; Q16388; Q16389;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Cartilage oligomeric matrix protein precursor (COMP).			
GN	COMP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cartilage;			
RX	MEDLINE=95229140; PubMed=7713493;			
RA	Newton G., Weremowicz S., Morton C.C., Copeland N.G.,			
RA	Gilbert D.J., Jenkins N.A., Lawler J.,			
RT	"Characterization of human and mouse cartilage oligomeric matrix			
RT	protein.";			
RL	Genomics 24:435-439(1994).			
[2]				
RP	VARIANTS PSACH SER-459 DEL; TYR-468 AND TYR-472.			
RX	MEDLINE=95400301; PubMed=7670471;			
RA	Hecht J.T., Nelson L.D., Crowder B., Wang Y., Elder F.F.B.,			
RA	Harrison W.R., Francomano C.A., Prange C.K., Lennon G.G., Deere M.,			
RA	Lawler J.;			
RT	"Mutations in exon 17B of cartilage oligomeric matrix protein (COMP)			
RT	cause pseudochondroplasia.";			
RL	Nat. Genet. 10:325-329(1995).			
[3]				
RP	VARIANTS MED TYR-342, AND VARIANT PSACH ARG-328.			
RX	MEDLINE=95400302; PubMed=7670472;			
RA	Briggs M.D., Hoffman S.M.G., King L.M., Oleen A.S., Mohrenweiser H.,			
RA	Leroy J.G., Mortier G.R., Rimoin D.L., Lachman R.S., Gaines E.S.,			
RA	Cekleniak J.A., Knowlton R.G., Cohn D.H.;			
RT	"Pseudochondroplasia and multiple epiphyseal dysplasia due to			
RT	mutations in the cartilage oligomeric matrix protein gene.";			
RL	Nat. Genet. 10:330-336(1995).			
[4]				
RP	VARIANT MED LYS-523.			

RX MEDLINE=97173141; PubMed=9021009;
 RA Ballejo R., Briggs M.D., Cohn D.H., Knowlton R.G., Beighton P.H.,
 RA Ramesar R.S.;
 RT "Multiple epiphyseal dysplasia, ribbing type: a novel point mutation
 RT in the COMP gene in a South African family.";
 RL Am. J. Med. Genet. 68:396-400(1997).
 RN [5]
 RP VARIANT MED SER-371, AND VARIANT PSACH 513-VAL--LYS-516 DEL.
 RX MEDLINE=97327574; PubMed=9184241;
 RA Susic S., McGrovy J., Ahler J., Cole W.G.;
 RT "Multiple epiphyseal dysplasia and pseudochondroplasia due to novel
 RT mutations in the calmodulin-like repeats of cartilage oligomeric
 RT matrix protein.";
 RL Clin. Genet. 51:219-224(1997).
 RN [6]
 RP VARIANTS PSACH AND MED.
 RX MEDLINE=98130533; PubMed=9463320;
 RA Briggs M.D., Mortier G.R., Cole W.G., King L.M., Golik S.S.,
 RA Bonneventure J., Nuytink L., de Paepe A., Leroy J.G., Blesacker L.,
 RA Lipson M., Wilcox W.R., Lachman R.S., Rimoin D.L., Knowlton R.G.,
 RA Cohn D.H.;
 RT "Diverse mutations in the gene for cartilage oligomeric matrix protein
 RT in the pseudochondroplasia-multiple epiphyseal dysplasia disease
 RT spectrum.";
 RL Am. J. Hum. Genet. 62:311-319(1998).
 RN [7]
 RP VARIANTS PSACH AND MED.
 RX MEDLINE=99118868; PubMed=9921895;
 RA Ikegawa S., Ohashi H., Nishimura G., Kim K.C., Sannohe A.,
 RA Kimizuka M., Fukushima Y., Nagai T., Nakamura Y.;
 RT "Novel and recurrent COMP (cartilage oligomeric matrix protein)
 RT mutations in pseudochondroplasia and multiple epiphyseal dysplasia.";
 RL Hum. Genet. 103:633-638(1998).
 RN [8]
 RP VARIANTS PSACH AND MED.
 RX MEDLINE=98112405; PubMed=9452026;
 RA Loughlin J., Irtan C., Mustafa Z., Briggs M.D., Carr A., Lynch S.-A.,
 RA Knowlton R.G., Cohn D.H., Sykes B.;
 RT "Identification of five novel mutations in cartilage oligomeric
 RT matrix protein gene in pseudochondroplasia and multiple epiphyseal
 RT dysplasia";
 RL Hum. Mutat. Suppl. 1:S10-S17(1998).
 RN [9]
 RP VARIANT PSACH GLY-482.
 RX MEDLINE=98112442; PubMed=9452063;
 RA Susic S., Ahler J., Cole W.G.;
 RT "Pseudochondroplasia due to the substitution of the highly conserved
 RT Asp82 by Gly in the seventh calmodulin-like repeat of cartilage
 RT oligomeric matrix protein.";
 RL Hum. Mutat. Suppl. 1:S15-S127(1998).
 CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.
 CC -1- DISEASE: DEFECTS IN COMP ARE THE CAUSE OF PSEUDOCHONDROPLASIA
 CC (PSACH) AND MULTIPLE EPIPHYSEAL DYSPLASIA (MED OR EDM1), WHICH ARE
 CC POSITIVELY INHERITED CHONDRODYSPLASIAS CHARACTERIZED BY SHORT
 CC STATURE AND EARLY-ONSET OSTEOARTHRITIS. MED IS BROADLY CATEGORIZED
 CC INTO THE MORE SEVERE FAIRBANK AND THE Milder RIBBING TYPES. PSACH
 CC IS MORE SEVERE AND IS RECOGNIZED IN EARLY CHILDHOOD.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: Contains 4 EGF-like domains.
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC -----
 DR EMBL, L32137; AAAS7253.1; -
 DR EMBL, S79499; AAB35289.1; -
 DR EMBL, S79500; AAB35270.1; -
 DR HGSP, P35444; 1VDF.

DR	Genew;	HGNC:2227;	COMP.	
DR	MIM;	600310;	-.	
DR	MIM;	132400;	-.	
DR	MIM;	177170;	-.	
DR	GO;	GO:0005578;	C:extracellular matrix; TAS.	
DR	GO;	GO:0005509;	F:calcium ion binding activity; TAS.	
DR	GO;	GO:0005201;	F:extracellular matrix structural constituent; TAS.	
DR	GO;	GO:0007397;	P:histogenesis and organogenesis; TAS.	
DR	GO;	GO:0001501;	P:skeletal development; TAS.	
DR	InterPro;	IPR001881;	EGF_Ca.	
DR	InterPro;	IPR006209;	EGF_like.	
DR	InterPro;	IPR003367;	EGF_3.	
DR	Pfam;	PF00008;	EGF; 2.	
DR	Pfam;	PF02412;	tsp_3; 11.	
DR	SMART;	SMO0179;	EGF_Ca; 2.	
DR	PROSITE;	PS00022;	EGF_1; FALSE_NEG.	
DR	PROSITE;	PS01166;	EGF_2; 1.	
DR	PROSITE;	PS01187;	EGF_Ca; 2.	
KW	Glycoprotein;	Cell adhesion;	Calcium-binding; Repeat; EGF-like domain;	
KW	Signal;	Disease mutation.		
FT	SIGNAL	1	20	POTENTIAL.
FT	CHAIN	21	757	CARTILAGE OLIGOMERIC MATRIX PROTEIN.
FT	DOMAIN	22	86	N-TERMINAL.
FT	DOMAIN	87	126	EGF-LIKE 1.
FT	DOMAIN	127	179	EGF-LIKE 2.
FT	DOMAIN	180	222	EGF-LIKE 3.
FT	DOMAIN	225	267	EGF-LIKE 4.
FT	DOMAIN	297	332	TSP TYPE-3 1.
FT	DOMAIN	333	355	TSP TYPE-3 2.
FT	DOMAIN	356	391	TSP TYPE-3 3.
FT	DOMAIN	392	414	TSP TYPE-3 4.
FT	DOMAIN	415	452	TSP TYPE-3 5.
FT	DOMAIN	453	488	TSP TYPE-3 6.
FT	DOMAIN	489	524	TSP TYPE-3 7.
FT	DOMAIN	525	757	C-TERMINAL.
FT	DISULFID	69	69	INTERCHAIN (PROBABLE).
FT	DISULFID	72	72	INTERCHAIN (PROBABLE).
FT	DISULFID	91	102	BY SIMILARITY.
FT	DISULFID	96	111	BY SIMILARITY.
FT	DISULFID	114	125	BY SIMILARITY.
FT	DISULFID	131	142	BY SIMILARITY.
FT	DISULFID	136	151	BY SIMILARITY.
FT	DISULFID	154	178	BY SIMILARITY.
FT	DISULFID	184	197	BY SIMILARITY.
FT	DISULFID	191	206	BY SIMILARITY.
FT	DISULFID	209	221	BY SIMILARITY.
FT	DISULFID	229	243	BY SIMILARITY.
FT	DISULFID	237	253	BY SIMILARITY.
FT	DISULFID	255	265	BY SIMILARITY.
FT	DISULFID	282	287	BY SIMILARITY.
FT	DISULFID	292	312	BY SIMILARITY.
FT	DISULFID	328	348	BY SIMILARITY.
FT	DISULFID	351	371	BY SIMILARITY.
FT	DISULFID	387	407	BY SIMILARITY.
FT	DISULFID	410	430	BY SIMILARITY.
FT	DISULFID	448	468	BY SIMILARITY.
FT	DISULFID	484	504	BY SIMILARITY.
FT	DISULFID	520	741	BY SIMILARITY.
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	742	742	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	367	369	CELL ATTACHMENT SITE (POTENTIAL).
FT	VARIANT	290	290	D -> N (IN PSAC; MILD FORM).
FT	VARIANT	299	299	/FTID=VAR 007614.
FT	VARIANT	299	299	G -> R (IN PSAC; MILD FORM).
FT	VARIANT	328	328	/FTID=VAR 007615.
FT	VARIANT	328	328	C -> R (IN PSAC; MILD FORM).
FT	VARIANT	342	342	/FTID=VAR 007616.
FT	VARIANT	342	342	D -> Y (IN MED; FAIRBANK TYPE).
FT	VARIANT	349	349	/FTID=VAR 007617.
FT	VARIANT	349	349	D -> V (IN PSAC; MILD FORM).
FT	VARIANT	349	349	/FTID=VAR 007618.
FT	VARIANT	361	361	D -> V (IN MED; FAIRBANK TYPE).
FT	VARIANT	361	361	/FTID=VAR 007619.

```
FT VARIANT 361 361 D -> Y (IN MED).
FT /FTID=VAR_007620.
FT VARIANT 367 368 MISSING (IN MED).
FT /FTID=VAR_007621.
FT VARIANT 371 371 C -> S (IN MED) FAIRBANK TYPE.
FT /FTID=VAR_007622.
FT VARIANT 372 372 MISSING (IN PSACH).
FT /FTID=VAR_007623.
FT VARIANT 374 374 MISSING (IN PSACH; MILD FORM).
FT /FTID=VAR_007624.
FT VARIANT 387 387 C -> G (IN PSACH; MILD FORM).
FT /FTID=VAR_007625.
FT VARIANT 391 394 PMSD -> V (IN PSACH).
FT /FTID=VAR_007626.

Alignment Scores:
Pred. No.: 2.1e-14
Score: 262.50 Length: 757
Percent Similarity: 75.61% Matches: 57
Best Local Similarity: 69.51% Mismatches: 9
Query Match: 39.06% Indels: 11
DB: 1 Gaps: 2

US-09-696-872-24 (1-372) x COMP_HUMAN (1-757)

OY 34 TTGCTCGCCCTTGCGCAGCTGTCAGCCGTCGCAAAAAGATCCAGCGTGGTGAGAC 93
Db 10 LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyGlnGlnSerProLeuGly----- 27
OY 94 TGTGTTGAGACCTGGGCGCCGAGATGCTTCGGGAACTGACGAAACCAACGCGCGCTG 153
Db 28 -----SerAspLeuGlyProGlnMetLeuArgIleuGlnGlnThrAsnAlaLeu 45
OY 154 CAGGACGTCGGGAGCTGCTGGCGACGACGATGACGAGGAGATGATCTCTGAAAAACG 213
Db 46 GlnAspValAlaGAspTrpLeuArgIlnGlnValArgIlnIleThrPheLeuLysAsnThr 65
OY 214 GTGATGAGTGTGACGCGCGCGCGCGAG-----CCGACGCG 252
Db 66 ValMetGluCysAspAlaCysGlyMetGlnGlnSerValArgThrGlyLeuProSerVal 85
OY 253 AAACCG 258
Db 86 ArgPro 87

RESULT 2
COMP_RAT STANDARD; PRT; 755 AA.
AC P35444;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cartilage oligomeric matrix protein precursor (COMP).
GN COMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RN SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93054522; PubMed=1429587;
RX Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;
RT "COMP (cartilage oligomeric matrix protein) is structurally related
RL to the thrombospondins.";
RL J. Biol. Chem. 267:22346-22350(1992).
RN (2)
RN X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 27-72.
RX MEDLINE=97020114; PubMed=8864111;
RX Malashkevich V.N., Kammerer R.A., Efimov V.P., Schultze T.,
RA Engel J.;
RT "The crystal structure of a five-stranded coiled coil in COMP: a
RT prototype ion channel?";
```

```
RL Science 274:761-765(1996).
CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72914; CA51419.1; -.
DR PIR: A44315; A44315.
DR PDB: 1VDF; 08-OCT-97.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR006209; EGF_Like.
DR InterPro: IPR003367; tsp_3.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF02412; tsp_3; 9.
DR SMART: SM00179; EGF_CA; 2.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 2.
KM Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
KW Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 755
FT DOMAIN 21 84
FT DOMAIN 85 124
FT DOMAIN 125 177
FT DOMAIN 178 220
FT DOMAIN 223 265
FT DOMAIN 295 330
FT DOMAIN 331 353
FT DOMAIN 354 389
FT DOMAIN 390 412
FT DOMAIN 413 450
FT DOMAIN 451 486
FT DOMAIN 487 522
FT DOMAIN 523 755
FT DISULFID 68 68
FT DISULFID 71 71
FT DISULFID 89 100
FT DISULFID 94 109
FT DISULFID 112 123
FT DISULFID 129 140
FT DISULFID 134 149
FT DISULFID 152 176
FT DISULFID 182 195
FT DISULFID 189 204
FT DISULFID 207 219
FT DISULFID 227 241
FT DISULFID 235 251
FT DISULFID 253 264
FT DISULFID 280 285
FT DISULFID 290 310
FT DISULFID 326 346
FT DISULFID 349 369
FT DISULFID 385 405
FT DISULFID 408 428
FT DISULFID 446 466
FT DISULFID 482 502
FT DISULFID 518 739
FT CARBOHYD 119 119
FT CARBOHYD 740 740
FT HELIX 30 66
FT TURN 67 67
FT HELIX 69 71
FT HELIX 71 71
SQ SEQUENCE 755 AA; 82663 MW; AB48888FE093C598 CRC64;
```

Alignment Scores:		4,76e-12	Length:	755
Pred. No.:		234.00	Matches:	51
Score:		72.84%	Conservative:	8
Percent Similarity:		62.96%	Mismatch:	18
Best Local Similarity:		34.82%	Indels:	4
Query Match:			Gaps:	1
DB:				

US-09-696-872-24 (1-372) x COMP_RAT (1-755)	
QY	34 TTGCTGCGCCCTTTGCGGAGCTGTGACGCGCTGCCAAAAAGATTCAGCGTGGTGAGAC 93
DB	9 ValLeuAlaLeuAlaAlaLeuArgAlaThrGlyGlnGlyGlnIleProLeuGlyGly--- 27
QY	94 TGTGTTCCAGACCTTGCGGCCCGGAGATGCTTGGGAACTCGAGAAACCAAGCGCGCTG 153
DB	28 -----AapLeuAlaProGlnMetLeuArgGlnLeuGlnGlnIleThrAspAlaLeu 44
QY	154 CAGAGCTGCGGAGACTGTGCTGCGGACAGGTCAGAGATCAGTCTCTGAAACACG 213
DB	45 GlnApValAlaArgGlnLeuLeuArgHisArgValIleGlnIleThrPheLeuArgAsnThr 64
QY	214 GTGATGAGGTGTGACGCGCTGCGGCGCGGACCGCAGCCGAAACCGCAGCCGACCGCGAG 273
DB	65 ValMetGlnGlyAspAlaCysGlyMetGlnProAlaArgThrProGlyLeuSerValArg 84
QY	274 CCG 276
DB	85 Pro 85

RESULT 3	
TSP4_XENLA	STANDARD; PRT; 955 AA.
AC	Q06441.
DT	01-JUN-1994 (Rel. 29, Created)
DT	01-JUN-1994 (Rel. 29, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Thrombospondin 4 precursor.
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxID=8355;
RP	[1]
RA	SEQUENCE FROM N.A.
EX	MEDLINE=93163109; PubMed=8432726;
RA	Lawler J., Duguet M., Whitaker C.A., Adams J.C., McHenry K.,
RT	Leishman D.W.;
RT	"Identification and characterization of thrombospondin-4, a new
RT	member of the thrombospondin gene family.";
RL	J. Cell Biol. 120:1059-1067(1993).
CC	- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC	LAMININ AND TYPE V COLLAGEN. MAY PARTICIPATE IN THE GENESIS AND
CC	FUNCTION OF CARDIAC AND SKELETAL MUSCLE.
CC	- SUBUNIT: Homotrimer; disulfide-linked.
CC	- DEVELOPMENTAL STAGE: INITIAL EXPRESSION DURING NEURULATION.
CC	INCREASE DURING TAILBUD STAGES BUT DECREASE BY THE PEDDING TADPOLE
CC	STAGE.
CC	- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC	- SIMILARITY: Contains 4 EGF-like domains.
CC	- SIMILARITY: Contains 7 TSP type-3 domains.
CC	- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC	or send an email to license@ebi.ac.uk).
CC	-----

[illegible]


```

Db      258  CysGlnAlaCysGlyProIleuLysPheGlnSerProThrProSer-ThrValValAlaPro 277
Oy      271  CAGCCGACGCCGAACCCGACGCCGAACCG 300
Db      278  AlaPro---ProAlaProProThrArgPro 286

RESULT 5
TSP4_RAT
ID_TSP4_RAT      STANDARD;      PRT;      980 AA.
AC P49744;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thrombospondin 4 precursor.
GN THBS4 OR TSP4 OR TSP-4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lewis; TISSUE=Skeletal muscle;
RX MEDLINE=96074771; PubMed=7490284;
RA Arder S., Caroni P.;
RT "Thrombospondin-4, an extracellular matrix protein expressed in the
RT developing and adult nervous system promotes neurite outgrowth.";
RL J. Cell Biol. 133:1083-1094 (1995).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----
DR EMBL; X89963; CAA62002.1; -.
DR HSSP; P35444; 1VDF.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR003567; tsp_3.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF000068; EGF; 2.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
KW Signal.
KW SIGNAL.
FT CHAIN 1 39 POTENTIAL.
FT CHAIN 40 980 THROMBOSPONDIN 4.
FT DOMAIN 40 303 N-TERMINAL.
FT DOMAIN 304 343 EGF-LIKE 1.
FT DOMAIN 344 396 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 397 437 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 438 481 EGF-LIKE 4.
FT DOMAIN 511 546 TSP TYPE-3 1.
FT DOMAIN 547 569 TSP TYPE-3 2.
FT DOMAIN 570 605 TSP TYPE-3 3.
FT DOMAIN 606 628 TSP TYPE-3 4.

```

FT	DOMAIN	629	666	TSP TYPE-3 . 5.
FT	DOMAIN	657	706	TSP TYPE-3 . 6.
FT	DOMAIN	707	742	TSP TYPE-3 . 7.
FT	DOMAIN	743	980	C-TERMINAL.
FT	DISUPEID	276	276	INTERCHAIN (PROBABLE).
FT	DISUPEID	279	279	INTERCHAIN (PROBABLE).
FT	DISUPEID	308	319	BY SIMILARITY.
FT	DISUPEID	313	328	BY SIMILARITY.
FT	DISUPEID	331	342	BY SIMILARITY.
FT	DISUPEID	348	359	BY SIMILARITY.
FT	DISUPEID	353	368	BY SIMILARITY.
FT	DISUPEID	371	395	BY SIMILARITY.
FT	DISUPEID	401	412	BY SIMILARITY.
FT	DISUPEID	406	421	BY SIMILARITY.
FT	DISUPEID	424	436	BY SIMILARITY.
FT	DISUPEID	442	456	BY SIMILARITY.
FT	DISUPEID	450	466	BY SIMILARITY.
FT	DISUPEID	468	480	BY SIMILARITY.
FT	DISUPEID	496	501	BY SIMILARITY.
FT	DISUPEID	506	526	BY SIMILARITY.
FT	DISUPEID	542	562	BY SIMILARITY.
FT	DISUPEID	565	585	BY SIMILARITY.
FT	DISUPEID	601	621	BY SIMILARITY.
FT	DISUPEID	624	644	BY SIMILARITY.
FT	DISUPEID	662	682	BY SIMILARITY.
FT	DISUPEID	708	722	BY SIMILARITY.
FT	DISUPEID	732	959	BY SIMILARITY.
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	960	960	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	980 AA;	108213 MM;	056D14EBE206FCE CRC64;

Alignment Scores:	
Pred. No.:	0.000119
Score:	144.50
Percent Similarity:	56.52%
Best Local Similarity:	42.03%
Query Match:	21.50%
DB:	1
Length:	960
Matches:	29
Conservative:	10
Mismatches:	27
Indels:	3
Gaps:	1

US-09-696-872-24 (1-372) X TSP4_RAT (1-980)

QY	102	GACCTGGGCGCCCAATGTCTTGGGAATCGCAGGAACAACGCGGGCGTCGCAGACGTG	163
Db	236	AAPPhenAsnArgGlnPheIeuLeuLysImethThrGlnIleuSerGlnLeuLeuGlyVal	255
QY	163	CGGAGCTGGCTGCGGCAGCAGGTTCAGGAGAATCATGTTCTGAAAACACCGGTATGAG	222
Db	256	LysAspLeuLeuArgGlnGlnValLysGlnIuhTrserPheLeuArgXanthrIleAlaGlu	275
QY	223	TGTGACGCGTGCGGGCCGACGCCGACCGCAAACCGCACCGCGCAG-----CGGCAG	273
Db	276	CySGlnAlaCyGlyProLeuSerPheGlnSerProThrProAdanThrLeuValProIle	295
QY	274	CGCAGCGCAAAACCGACGCCAACCAGCG300	
Db	296	AlaProIleAlaProProThrArgPro304	

CSP_PLABE	STANDARD;	PRT;	339 AA.
ID _CSP_PLABE			
AC	P06915;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	01-JAN-1988 (Rel. 06, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Circumsporozoite protein precursor (CS).		
OS	Plasmodium berghei.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_Taxid=5821;		
RN	[1] _SEQUENCE FROM N.A.		
RP	MEDLINE=87089740; PubMed=2432195;		
RX	Etchinger D.J., Arnot D.E., Tam J.P., Nussenzweig V., Enea V.;		
RT	"Circumsporozoite protein of Plasmodium berghei: gene cloning and		

RT identification of the immunodominant epitopes.",
 RL Mol. Cell. Biol. 6:3965-3972(1986).
 CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
 CC VERTEBRATE HOST).
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
 CC -1- SIMILARITY: Contains 1 TSP type-1 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14135; AAA29577.1; -.
 CC PIR: A44948; OZZOMB.
 CC InterPro: IPR003067; Circmsprzoite.
 CC InterPro: IPR000884; TSP1.
 CC Pfam: PF00090; tsp_1; 1.
 CC PRINTS: PR01303; CRCMSPRZOITE.
 CC SMART: SM00209; TSP1.1.
 CC PROSITE: PS50092; TSP1.1.
 CC Malaria; Sporozoite; Repeat; Signal.
 CC SIGNAL 1 23 PROBABLE.
 CC CHAIN 24 339 CIRCUMSPOROZOITE PROTEIN.
 CC DOMAIN 93 196 13 X 8 AA TANDEN REPEATS.
 CC DOMAIN 206 238 16 X 2 AA TANDEN REPEATS OF P-Q.
 CC DOMAIN 266 317 TSP TYPE-1.
 CC SEQUENCE 339 AA; 37138 MW; E8068A6D11D9551B CRC64;
 SO
 Alignment Scores:
 Pred. No.: 0.00109 Length: 339
 Score: 133.00 Matches: 20
 Percent Similarity: 87.50% Conservative: 8
 Best Local Similarity: 62.50% Mismatches: 4
 Query Match: 19.79% Indels: 0
 Gaps: 0
 DB: 1
 US-09-696-872-24 (1-372) x CSP_PLABE (1-339)
 QY 238 CCGGAGCCGCGAGCCGGAACCGGAGCCGCGAGCCGCGGAACCGGAGCCGGA 297
 DB 216 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 235
 QY 298 CCGGAACCGGAAGGTACCGGATCATCAGAAAAAGAT 333
 DB 236 ProGlnProGlyGlyAaAsnAsnAsnAsnAsn 247
 RESULT 7
 ID CSP_PLABA STANDARD; PRT; 347 AA.
 AC P23093;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Circumsporoite protein precursor (CS).
 OS Plasmodium berghei (strain Anka).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 CC NCBI_TaxID=5823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90221834; PubMed=2183186;
 RA Lockyer M.J., Davies C.S., Sunbrier A., Sinden R.E.;
 RT "Nucleotide sequence of the Plasmodium berghei circumsporoite
 RT protein gene from the ANKA clone 2.34L."
 RL Nucleic Acids Res. 18:376-376(1990).
 CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE

CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
 CC VERTEBRATE HOST).
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
 CC -1- SIMILARITY: Contains 1 TSP type-1 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X17606; CA35608.1; -.
 CC PIR: S07873; OZZOBK.
 CC InterPro: IPR003067; Circmsprzoite.
 CC InterPro: IPR000884; TSP1.
 CC Pfam: PF00090; tsp_1; 1.
 CC PRINTS: PR01303; CRCMSPRZOITE.
 CC SMART: SM00209; TSP1.1.
 CC PROSITE: PS50092; TSP1.1.
 CC Malaria; Sporozoite; Repeat; Signal.
 CC SIGNAL 1 23 POTENTIAL.
 CC CHAIN 24 347 CIRCUMSPOROZOITE PROTEIN.
 CC DOMAIN 93 204 13 X 8 AA REPEATS.
 CC DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.
 CC DOMAIN 274 325 TSP TYPE-1.
 CC SEQUENCE 347 AA; 37776 MW; 0EC240EE35681AF8 CRC64;
 SO
 Alignment Scores:
 Pred. No.: 0.00108 Length: 347
 Score: 133.00 Matches: 20
 Percent Similarity: 87.50% Conservative: 8
 Best Local Similarity: 62.50% Mismatches: 4
 Query Match: 19.79% Indels: 0
 Gaps: 0
 DB: 1
 US-09-696-872-24 (1-372) x CSP_PLABA (1-347)
 QY 238 CCGGAGCCGCGAGCCGGAACCGGAGCCGCGAGCCGCGGAACCGGAGCCGGA 297
 DB 224 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 243
 QY 298 CCGGAACCGGAAGGTACCGGATCATCAGAAAAAGAT 333
 DB 244 ProGlnProGlyGlyAaAsnAsnAsnAsnAsn 255
 RESULT 8
 ID TEGU_HSV11 STANDARD; PRT; 3164 AA.
 AC P10220;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Large tegument protein (Vinton protein UL36).
 OS UL36.
 OS Herpes simplex virus (type 1 / strain 17).
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Alphaherpesvirinae; Simplexvirus.
 CC NCBI_TaxID=10299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88274327; PubMed=2839594;
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;
 RT "The complete DNA sequence of the long unique region in the genome of
 RT herpes simplex virus type 1."
 RL J. Gen. Virol. 69:1531-1574(1988).
 CC -1- FUNCTION: TEGUMENT PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC EBV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.

```

CC -----
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, X14112, CAA32311.1; ..
CC DR PIR, I30085; WMBE6H.
CC DR InterPro, IPR006928; Herpes teg N.
CC DR InterPro, IPR005210; Herpes_UL36.
CC DR Pfam, PF04843; Herpes teg N/1.
CC DR Pfam, PF03586; Herpes_UL36/1.
CC Repeat.
CC KW Repeat.
CC FT DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
CC SEQUENCE 3164 AA; 335857 MW; CCS5D31FF49F9E3F4 CRC64;
CC -----
CC Alignment Scores:
CC Pred. No.: 0.00126 Length: 3164
CC Score: 132.00 Matches: 22
CC Percent Similarity: 81.82% Conservative: 5
CC Best Local Similarity: 66.67% Mismatches: 6
CC Query Match: 19.64% Indels: 0
CC DB: 1 Gaps: 0
CC -----
US-09-696-872-24 (1-372) x TEGU HSV11 (1-3164)
QY 211 ACGGTGATGTGACGTGTGACGCGTGGCGGGCGAGCGAGCGGAAACCGCAGCGAGCGG 270
Db 2904 ThrValSerArgLeuSerHisAlaProGlnProGlnProGlnProGlnProGlnPro 2923
QY 271 CAGCCGAGCCGAAACCGCAGCGAGCGGAAACCGGAGACCGGAA 309
Db 2924 GlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 2936
-----
RESULT 9
NOL3_MOUSE STANDARD; PRT; 220 AA.
AC Q9D1X0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleolar protein 3.
GN NOL3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=adipose tissue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batlora S., Casavant T.,
RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Peele G., Quackenbush J.,
RA Schirral L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guestrich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Matzarelli U., Mombauts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Winding L.,
RA Wyshewski-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."

```

```

RL Nature 409:685-690(2001).
CC -!- FUNCTION: May be involved in RNA splicing (By similarity).
CC -!- SUBUNIT: Interacts with Srp30c, NPM1, CASP2, CASP8 and CED-3 (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 CARD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AK021023; BAB32281.1; -.
DR MCD; MGI:1925938; N013.
DR InterPro; IPR001315; CARD.
DR SMART; SM0114; CARD; 1.
DR PROSITE; PS50209; CARD; 1.
DR KMW Nuclear protein; mRNA splicing.
FT DOMAIN 4 95
FT DOMAIN 132 218 GLU/PRO-RICH.
SQ SEQUENCE 220 AA; 24567 MW; AADCD57C1EB320A2 CRC64;

Alignment Scores:
Pred. No.: 0.00607 Length: 220
Score: 124.00 Matches: 44
Percent Similarity: 41.33% Conservative: 18
Best Local Similarity: 29.33% Mismatches: 36
Query Match: 18.45% Indels: 52
DB: 1 Gaps: 7

US-09-696-872-24 (1-372) x NOL3_MOUSE (1-220)
QY 32 GCTTGCTGCGCCCTG-----CGGAGCTGTCAGCGCTGCCAAAAAGATGCA 79
   ||||| ||| ::|||::: |||||
DB 72 AIAcySgInGluLLeuAArgCyAlGInGlnThValAlargwEPro-----AspPro 89
   |||||
QY 80 GCTTGCGTGAGAGACTGTTGTTCAAGCCTGGG----- 110
   |||||
DB 90 AIAtrPaeRTrGInLInLVal-GLYProGILTYTrAlGAsnAArgSerTYrAaProSerCY 109
   |||||
QY 111 -----CCCGAGATGCTTCCGGGAAC----- 131
   |||||
DB 109 sProGILYHieRTrPnRProGILaIaProSerSerGILThRThCySProGILLeuProAr 129
   |||||
QY 132 -GCGAGAAACCAACGCGCGCTGCGAGACGTGGGGA----- 167
   |||||
DB 129 gAlaSerGInGInGInGInuAlGlyGlyProGInGILSerGILaLeuGInProArGTh 149
   |||||
QY 168 -----CTTGCTGCGGCGAGCGAGTCAAGGAGATCAAGT----- 200
   |||||
DB 149 rProGInGInuProGILLeuGInuAlaGInuAlaThGInGILYAspGILuProAspLeuGInG 169
   |||||
QY 201 -----CTTGAAAAACCGGTGATGAGTGTGACCGCTGGCGGCGCGAGCGCAGC 250
   |||||
DB 169 nGILuMeFAsnProGInGInuProGInuProGILuProGILuProGILuProGILuProGILu 189
   |||||
QY 251 CGAAACCGGAGCGCGAGCCGAGCCGAGCCGAAACCGGAGCCGAAACCGGAAACCGGAA 310
   |||||
DB 189 rOGILuPProGILuProGILuProGILuProGILuProGILuProGILuProGILuProGILu 208
   |||||
QY 311 GTACCGGATCATCAGAAAAAGATAG 336
   |||||
DB 209 --ProAspPneGInGInGILuAspGILu 216
   |||||

RESULT 10
ID_SFRA_MOUSE STANDARD; PRT; 283 AA.
AC Q60701.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Splicing factor, arginine/serine-rich 10 (Putative myelin regulatory
 factor 1) (MRF-1) (fragment).
 GN SFRS10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95074156; PubMed=7527040;
 RA Haque N.S., Buchberg A.M., Khalili K.;
 RT "Isolation and characterization of MRF-1, a brain-derived DNA-binding
 protein with a capacity to regulate expression of myelin basic protein
 gene."
 RL J. Biol. Chem. 269:31149-31156(1994).
 CC -1- FUNCTION: CAN BIND TO THE MYELIN BASIC PROTEIN (MBP) GENE MB3
 CC REGULATORY REGION AND INCREASE TRANSCRIPTION OF THE MBP PROMOTER
 CC IN CELLS DERIVED FROM THE CNS.
 CC -1- SUBCELLULAR LOCATION: Nucleus (By similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL THE TISSUES EXAMINED (LIVER,
 CC KIDNEY, SPLEEN, HEART, LUNG, AND BRAIN).
 CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
 CC DOMAIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U14648; AAA64595.1; -.
 DR MGD; MGI:104846; Sfrs10.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 KW Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation;
 KW DNA-binding.
 FT NON_TER 1 1
 FT DOMAIN 167 172 GLY-RICH (HINGE REGION).
 FT DOMAIN 173 259 ARG/SER-RICH (RS DOMAIN).
 SQ SEQUENCE 283 AA; 31581 MW; 084EB2C9CC9172AE CRC64;
 Alignment Scores:
 Pred. No.: 0.00884 Length: 283
 Score: 122.00 Matches: 42
 Percent Similarity: 57.01% Conservative: 19
 Best Local Similarity: 39.25% Mismatches: 22
 Query Match: 18.15% Indels: 24
 DB: 1 Gaps: 6
 US-09-696-872-24 (1-372) x SFR_MOUSE (1-283)
 Oy 38 TCGCCCTGGCGGCACTGCGAGCGCTGCCAAAAGATCCAGCTGGGTGAGACTGTT 97
 Db 176 SerProAAGAGTProAlaAlaIleSerlySProGlu----- 188
 Oy 98 GTTCAGACTGGGCGCGGAGATCTTCGGAACCAACGCGCGCTGCAGG 157
 Db 189 -----TtpSerArg-----ArglySProSerArglySerArg 199
 Oy 158 ACGTGGGAGTGGTGGCGGAGAGTTCAGGAGATCACTTCTGAAAACACGGTGA 217
 Db 200 Ser-----LySerArgSer--ArgTlnArgSer--ArgSerArgSer 212
 Oy 218 TGGAGTGTGAGCGCT--CGCGGCGCGAGCGCGACCAACCGAGCGCGAGCGCAGC 274
 Db 213 ThrSerLySerArgSerArglyArglySerlySerSerSerSerValSerArgSer 232
 Oy 275 CGACGCCAAACCGAGCGGAAACCGAAGCGAAGTACCGGATCATCAGAAAAGATG 334

Db 233 ArgSerArgSerArgSerArgSerArgSerArgSerProProValSerlyArgGlu 252
 Oy 335 AGTGTAGCGCGCGCGAGAT 355
 Db 253 SerlySerArgSerArgSer 259
 RESULT 11
 PARB_TRYB STANDARD; PRT; 129 AA.
 AC P09791;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, last sequence update)
 DE 16-OCT-2001 (Rel. 40, last annotation update)
 DE Procyelic form specific polypeptide A-beta precursor (Procyelin) (PARB
 DE A-beta).
 GN PARP-BETA.
 OS Trypanosoma brucei brucei.
 OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_Taxid=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=427;
 RX MEDLINE=87115776; PubMed=3808022;
 RA Roditi I., Carrington M., Turner M.;
 RT "Expression of a polypeptide containing a dipeptide repeat is
 RT confined to the insect stage of Trypanosoma brucei."
 RL Nature 325:272-274(1987).
 (2)
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90258895; PubMed=2342468;
 RX Clayton C.E., Frier J.P., Itzhaki J.E., Bellofatto V., Sherman D.R.,
 RA Wisdom G.S., Vijayasarathy S., Mowatt M.R.;
 RT "Transcription of the procyelic acidic repetitive protein genes of
 RT Trypanosoma brucei."
 RL Mol. Cell. Biol. 10:3036-3047(1990).
 (3)
 RN SEQUENCE FROM N.A.
 RP STRAIN=427;
 RA Vijayasarathy S., Ernest I., Itzhaki J., Sherman D., Mowatt M.R.,
 RA Michaels P.A.M., Clayton C.E.;
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 (4)
 RP SEQUENCE OF 26-64, AND POST-TRANSLATIONAL MODIFICATIONS.
 RC STRAIN=427;
 RX MEDLINE=89359323; PubMed=2475493;
 RA Clayton C.E., Mowatt M.R.;
 RT "The procyelic acidic repetitive proteins of Trypanosoma brucei.
 RT Purification and post-translational modification."
 RL J. Biol. Chem. 264:15088-15093(1989).
 CC -1- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
 CC DIFFERENTIATION IN THE INSECT VECTOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04814; CAA28503.1; -.
 DR EMBL; M33129; AAA30225.1; -.
 DR EMBL; X52584; CAA36815.1; -.
 DR PIR; A26036; A26036.
 KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.
 FT SIGNAL 1 27
 FT CHAIN 28 107
 FT PROPEP 108 129
 FT DOMAIN 59 106 24 X 2 AA TANDEM REPEATS OF [DE]-P-B.

```

FT CARBOHYD 55 55 O-LINKED.
FT LIPID 107 107 GPI-ANCHOR.
SQ SEQUENCE 129 AA; 13314 MW; 063A1E64D817B86 CRC64;

Alignment Scores:
Pred. No.: 0.0144 Length: 129
Score: 119.50 Matches: 25
Percent Similarity: 56.82% Conservative: 15
Best Local Similarity: 36.76% Mismatches: 16
Query Match: 17.78% Indels: 12
DB: 1 Gaps: 3

US-09-696-872-24 (1-372) x PARB_TRYBB (1-129)

QY 108 GGGCCCGCAGATGCTTCGGGAACCTGAGGAAACCAACGCGCGCTGACGAGCGGGA 167
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30 GYPRGGLUAP-----LysGlyLeuThrIlyGlyLysGlyLys----- 43
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 168 CTGGCTGCGGACGAGTCAGGAGATCAGTTCTGAAAAACGCGTGATGAGTGCA 227
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 44 -----GlyGlyLysGlyThrIlyValSerAspAspAspThrAsnGlyThr-As 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 228 CGCGTGGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 287
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 59 PPro--AspProGluProGluProGluProGluProGluProGluProGluProGluPr 78
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 288 GCAGCGGAAACCGGAAACCGGAA 309
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 cqluProGluProGluProGlu 85

RESULT 12
TONB_HELPY STANDARD; PRT; 285 AA.
ID TONB_HELPY AC 025899;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TONB protein.
GN TONB OR HPI341.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
CX NCBI_TaxID=210;
(1)
RN SEQUENCE FROM N.A.
RP STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA J-F., White O., Kersavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Melman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Frazer C.M.,
RA Venter J.C.;
RA "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -I- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AE000635; AAD08383.1; -.
CC PIR; E64687; E64687.
CC TIGR; HP1341; -.
CC InterPro; IPR003538; TONB.
CC InterPro; IPR006260; TONB_C.
CC Pfam; PF03544; TONB_1.
CC PRINTS; PR01374; TONBPROTEIN.
CC TRIPFAMS; TIGR01352; TONB_Cterm; 1.
CC Transport; Protein transport; Inner membrane; Periplasmic;
CC Transmembrane; Signal-anchor; Repeat; Complete proteome.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 35 SIGNAL-ANCHOR (POTENTIAL).
FT DOMAIN 36 285 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 285 AA; 31585 MW; 3897A8B7B70BDF8 CRC64;

Alignment Scores:
Pred. No.: 0.0229 Length: 285
Score: 117.00 Matches: 20
Percent Similarity: 79.49% Conservative: 11
Best Local Similarity: 51.28% Mismatches: 7
Query Match: 17.41% Indels: 1
DB: 1 Gaps: 0

US-09-696-872-24 (1-372) x TONB_HELPY (1-285)

QY 238 CCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106 ProLysProLysProLysProLysProLysProLysProLysProLysProLysProLys 125
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 298 CCGGACCGGAAAGTACCGGATCATAGAAAAAGTGTGTAGCGGCGCGGAC 352
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 ProGluPro-LysValGluGluValLysLysGluGluProLysGluGluProLys 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
PARB_TRYBB STANDARD; PRT; 115 AA.
ID PARB_TRYBB AC P14043;
AC 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Procyelic form specific polypeptide precursor (Procyelin) (PARP).
GN PROX.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
CX NCBI_TaxID=5702;
(1)
RN SEQUENCE FROM N.A.
RP STRAIN=227 (TILPAT);
RX MEDLINE=90067841; PubMed=2573878;
RA Koenig E., Delius H., Carrington M., Williams R.O., Roditi I.;
RA "Duplication and transcription of procyelin genes in Trypanosoma
RT brucei."
RL Nucleic Acids Res. 17:8727-8739(1989).
CC -I- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -I- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
CC DIFFERENTIATION IN THE INSECT VECTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X16015; CAA34147.1; -.

```


CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M3130; AAA30226.1; -
DR EMBL; X52585; CAA36816.1; -
DR EMBL; L02933; AAA30228.1; -
DR PIR; A4418; A4418.
KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.
FT SIGNAL 1 27 BY SIMILARITY.
FT CHAIN 28 123 PROCYCLIC FORM SPECIFIC POLYPEPTIDE B-
FT PROPEP 124 145 ALPHA.
FT DOMAIN 59 122 BY SIMILARITY.
FT LIPID 123 123 32 X 2 AA TANDEN REPEATS OF [DE]-P.
SQ SEQUENCE 145 AA; 15107 MW; B6C3B529BD794CE3 CRC64;

Alignment Scores:

Pred. No.:	0.0255	Length:	145
Score:	116.50	Matches:	22
Percent Similarity:	54.41%	Conservative:	15
Best Local Similarity:	32.35%	Mismatches:	17
Query Match:	17.34%	Indels:	14
DB:	1	Gaps:	1

US-09-696-872-24 (1-372) x PARC_TRYBB (1-145)

QY	108	GGGCGCGGAGATGCTTCGGGAACTGCGAGAACCAACGGCGGCTGCGAGACGTGCGGGA	167
DB	30	GlyProGluAspLysGlyLeuThrLysGlyLys-----	41
QY	168	CTGGCTGCGGAGAGATGCGAGATCACTTCTGAAAAACACGATGATGAGTGA	227
DB	42	-----GlyLysGlyLysGlyLysValGlyAlaAsp-AspThrAs	56
QY	228	CGCGTGGCGGCGGCGGAGCGGCGGAAACCGCGGCGGCGGAGCGGCGGAAAC	287
DB	56	nGlyThrAspProAspProGluProGluProGluProGluProGluProGluPro	76
QY	288	GCAGCGGAACCGGAACCGGA	309
DB	76	OGluProGluProGluProGlu	83

Search completed: August 26, 2003, 16:22:06
Job time : 21.5 secs

Alignment Scores:

Pred. No.:	1.65e-15	Length:	130
Score:	256.50	Matches:	8
Percent Similarity:	63.30%	Conservative:	61
Best Local Similarity:	55.96%	Mismatches:	23
Query Match:	38.17%	Indels:	17
DB:	4	Gaps:	3

US-09-696-872-24 (1-372) x Q8N2R4 (1-130)

QY	3	TTCTCGACCTTTCGGGCACTGCAGCCGTCGCAAAAAAGGATCCAGCTGGTGTGGAGAC	93
		: : : : : : : : : : : : : : : :	
Db	10	LeuLeuThrLeuAlaIleuGluYAlaSerGlyIleGlnGlyIleSerProLeuGlyI-	27
QY	94	TGTTTTCAGACCTGGGCCCCGAGATGCTTCCGGAACTGCAGAGAAACCAACGCGCGCGCTG	153
Db	28	-----SerIlePheuGlyProGlnIleLeuArgGluIleuGlnGluThrAsnAlaIleu	45
QY	154	CAGACAGTCGGGAGCTGCGTCGGCGCAGCAGTCAAGGAGATCACGTTCTTAATAAACAGC	213
Db	46	GlnAspValArgGluIleuLeuArgGlnGlnValArgGluIleThrPheLeuYAlaSerThr	65
QY	214	GTATGTGAGTGTACCCGCTGCGGCGCGCAGG-----CCGCACCG	252
Db	66	ValMetGluIcyAspAlaCysGlyMetGlnGlnSerValArgThrGlyLeuProSerVal	85
QY	253	AAACCGCAGCCGACGCGCCGACCG-----CAGCCGAAACCGCAGCCG	294
Db	86	ArgProLeuLeuIleIcyAlaIleProGlyValProLeuArgGlyLeuProAlaGlyValGln	105
QY	295	AAACCGGAACCGGAAGTACCGGATCA	321
Db	106	ArgProIleProProGlyArgGlyAla	114

RESULT 2

ID	Q9BG80	PRELIMINARY;	PRT;	755 AA.
AC	Q9BG80;			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Cartilage oligomeric matrix protein.			
GN	COMP.			
OS	Equus caballus (Horse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;			Euteleostomi;
OC	Mammalia; Eutheria; Perissodactyla; Equidae; Equus			
NCBI_TaxID=9796;				

Pred. No.:	1.21e-14	Length:	755
Score:	247.50	Matches:	55
Percent Similarity:	75.86%	Conservative:	11
Best Local Similarity:	63.22%	Mismatches:	16

Query Match:	36.83%	Indels:	5
DB:	6	Gaps:	2

US-09-696-872-24 (1-372) x Q9BG80 (1-755)

QY 3 TTGCTGCGCCCTTCCGAGATGCTGCGCGCTCCAAAAAAGGATCCACCTGGGTGGAGAC 93
Db 10 LeuLeuAlaLeuAlaAlaLeuValSerSerIlnIyctln---ThrProLeuGly----- 26
QY 94 TGTGTTTCAGACCTTGGGCGCCGAGATGCTTCGGAACTGCAGGAAACCAACGCGCGCTG 153
Db 27 -----ThiGlnLeuGlyProGlnMetLeuArgIlnLeuGlnGlnTrpAlaAlaLeu 44
QY 154 CAGGACGTGCGGAGCTGGCTGCGGAGAGGTTCAGGGAGATCACTGCTCGAAAAACAGC 213
Db 45 GlnAspValArgGlnLeuLeuArgGlnIlnAlaValSerIlnIleThrPheLeuIyAsnThr 64
QY 214 GTGATGAGATGTGACCGCGTGGCGGCGCGAGCCGCAAGCCGAAACCGACCGCGAG 273
Db 65 ValMetGlnCysAspAlaCysGlyMetGlnProAlaArgThrProArgValSerValArg 84
QY 274 CCGAGCCGAAACCGAGCGC 294
Db 85 ProLeuAlaGlnCysAlaPro 91

RESULT 3

ID	AC	014592;	PRELIMINARY;	PRT;	817 AA.
DT	DT	01-JAN-1998	(TRENBLER)	05,	Created
DT	DT	01-JAN-1998	(TRENBLER)	05,	Last sequence update)
DT	DT	01-MAR-2003	(TRENBLER)	23,	Last annotation update)
DE	DE	COMP_HUMAN.			
GN	GN	COMP			
OS	OS	Homo sapiens (Human).			
OC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	OX	NCBI_TaxID=9606;			
RN	RN	[1]			
RP	RP	SEQUENCE FROM N.A.			
RA	RA	Iamedin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,			
RA	RA	Gordon L., Kyle A., Ramlitz M., Stiwegen S., Ganes J., Dangnan			
RA	RA	Christensen M., Bruce R., Quan G., Montgomery M., Ow D., Kobayashi			
RA	RA	Olsen A.O., Carraro A.V.;			
RT	RT	"Sequence analysis of an ~1 Mb region containing the MER2B gene in			
RL	RL	1992."			
Submitted	Submitted	(NOV-1997) to the EMBL/GenBank/DBJ databases.			

Alignment Scores:

Pred. No.:	1.66e-14	Length:	817
Score:	245.50	Matches:	5
Percent Similarity:	74.39%	Conservative:	66
Best Local Similarity:	67.07%	Mismatches:	10
Query Match:	36.53%	Indels:	11
DB:	4	Gaps:	2

US-09-696-872-24 (1-372) x 014592 (1-817)

34 TTGCTGCCCCCTTCCGGCAGTCTGCAGCGCTGCCAAAAAAGATCCAGCCTGGGTGGAGAC 93

Db 10 LeuLeuThrLeuAlaAlaLeuGlyAlaIaSerGlyInGlyInserProLeuGly----- 27
 QY 94 TGTGTTAGACCTGGGCGCCGACAGTGTCTGGGGAACCTGCAGAAACCAAGCGCGCTG 153
 Db 28 -----SerAspLeuGlyProGlnMetLeuArgGlnLeuGlnThrAspAlaAlaLeu 45
 QY 154 CAGGACGTGCGGACCTGGCTGCGGACGAGGTCAAGGAGATCACCTTCCTGAAAAACAG 213
 Db 46 GlnAspValArgGlnLeuLeuArgGlnInValArgGlnIleThrPheLeuLysAsnThr 65
 QY 214 GTGATGAGTGTACGCGCTGGCGCGCAG-----CCGACGCG 252
 Db 66 ValMetGlnCysAspAlaCysGlyMetGlnInserValArgThrGlyLeuProSerVal 85
 QY 253 AACCG 258
 Db 86 ArgPro 87
 RESULT 4
 Q9R0G6 PRELIMINARY; PRT; 755 AA.
 AC Q9R0G6.
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cartilage oligomeric matrix protein precursor.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cartilage.
 RA Fang C., Carlson C.S., Leslie M.P., Tull H., Stoleran E., Perris R.,
 RA Ni L., Di Cesare P.E.;
 RT "Molecular Cloning, Sequencing, Tissue and Developmental Expression of
 RT Mouse Cartilage Oligomeric Matrix Protein (COMP).",
 RL J. Orthop. Res. 0:0-0(1999).
 DR EMBL; AF033530; AAD01972.1; -.
 DR HSSP; P35444; IVDP.
 DR MGD; MGI:88469; Comp.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR00367; tsp_3.
 DR Pfam; PF02412; tsp_3; 9.
 DR SMART; SM00179; EGF_CA; 2.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 2.
 KM EGF-like domain; Matrix protein; Signal.
 FT SIGNAL 1
 FT CHAIN 19
 FT POTENTIAL.
 SQ SEQUENCE 755 AA; 82352 MW; 7DDFCF443589A0B7 CRC64;

Alignment Scores:
 Pred. No.: 9.23e-14 Length: 755
 Score: 238.00 Matches: 54
 Percent Similarity: 66.32% Conservative: 9
 Best Local Similarity: 56.84% Mismatches: 28
 Query Match: 35.42% Indels: 4
 Gaps: 1
 DB: 11

US-09-696-872-24 (1-372) x Q9R0G6 (1-755)

QY 10 ATGGAAGTACATGATTAGGCTTGGCCCTTGGCGGAGTGTGAGCGCTGCCAA 69
 Db 1 MetGlyProThrAlaCysValLeuValLeuAlaAlaIleLeuArgAlaThrGlyIn 20
 QY 70 AAAGATCCAGCTGGGTGAGACTGTTGTCAGACTGGGCGCCAGATGCTTCGGGA 129
 Db 21 GlyInIleProLeuGlyGly-----AspLeuAlaProGlnMetLeuArgGln 36

QY 130 CTGAGGAAACCAAGCGGCGCTGCAGAGCTGTGCGGACGTGCGGACGAGTCCAG 189
 Db 37 LeuGlnGlnThrAlaAlaAlaLeuGlnAspValArgGlnLeuLysArgIleValLys 56
 QY 190 GAGATCAGTTCCTGAATAACACCGTGATGAGTGTGACCGCTGCGGCGCCAGCCGAG 249
 Db 57 GlnIleThrPheLeuLysAsnThrValMetGlnCysAspAlaCysGlyMetGlnProAla 76
 QY 250 CCGAAACCGCAGCCGCGACCCGACCCGCAACCGCAGCCG 294
 Db 77 ArgThrProGlyLeuSerValArgProValProLeuCysAlaPro 91
 RESULT 5
 Q8V154 PRELIMINARY; PRT; 755 AA.
 ID Q8V154.
 AC Q8V154.
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cartilage oligomeric matrix protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fang C., Leslie M.P., Nord R., Tian H., Di Cesare P.E.;
 RT "Mouse Cartilage Oligomeric Matrix Protein Genomic Sequence.",
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF257516; AAL36518.1; -.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR00367; tsp_3.
 DR Pfam; PF02412; tsp_3; 9.
 DR SMART; SM00179; EGF_CA; 2.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 2.
 KM EGF-like domain; Matrix protein.
 SQ SEQUENCE 755 AA; 82286 MW; 91758FCF789167EB CRC64;

Alignment Scores:
 Pred. No.: 9.23e-14 Length: 755
 Score: 238.00 Matches: 54
 Percent Similarity: 66.32% Conservative: 9
 Best Local Similarity: 56.84% Mismatches: 28
 Query Match: 35.42% Indels: 4
 Gaps: 1
 DB: 11

US-09-696-872-24 (1-372) x Q8V154 (1-755)

QY 10 ATGGAAGTACATGATTAGGCTTGGCCCTTGGCGGAGTGTGAGCGCTGCCAA 69
 Db 1 MetGlyProThrAlaCysValLeuValLeuAlaAlaIleLeuArgAlaThrGlyIn 20
 QY 70 AAAGATCCAGCTGGGTGAGACTGTTGTCAGACTGGGCGCCGAGATGCTTCGGGA 129
 Db 21 GlyInIleProLeuGlyGly-----AspLeuAlaProGlnMetLeuArgGln 36
 QY 130 CTGAGGAAACCAAGCGGCGCTGCAGAGCTGTGCGGACGTGCGGACGAGTCCAG 189
 Db 37 LeuGlnGlnThrAlaAlaAlaLeuGlnAspValArgGlnLeuLysArgIleValLys 56
 QY 190 GAGATCAGTTCCTGAATAACACCGTGATGAGTGTGACCGCTGCGGCGCCAGCCGAG 249
 Db 57 GlnIleThrPheLeuLysAsnThrValMetGlnCysAspAlaCysGlyMetGlnProAla 76
 QY 250 CCGAAACCGCAGCCGCGACCCGACCCGCAACCGCAGCCG 294
 Db 77 ArgThrProGlyLeuSerValArgProValProLeuCysAlaPro 91

```
RESULT 6
Q8N4T2 PRELIMINARY; PRT; 724 AA.
ID Q8N4T2
AC Q8N4T2
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to cartilage oligomeric matrix protein (pseudochondroprolasi,
DE epiphyseal dysplasia 1, multiple).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC Tissue=Brain, and Lung;
RA Strauberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033676; AAH33676.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR SMART; SM00181; EGF; 4.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW Matrix Protein.
SQ SEQUENCE 724 AA; 79696 MW; 9AE2DB3F88815FA1 CRC64;

Alignment Scores:
Pred. No.: 8.12e-10 Length: 724
Score: 195.50 Matches: 41
Percent Similarity: 79.63% Conservatve: 2
Best Local Similarity: 75.93% Mismatches: 4
Query Match: 29.09% Indels: 7
DB: Gaps: 1

US-09-696-872-24 (1-372) x Q8N4T2 (1-724)
QY 118 ATGCTTCGGGAACCTGCGAAGAAACCAACGCGCGCTGCGAGAGCTGGGACTGCTGGCG 177
DB 1 MetLeuAAGGluLeuGlnGlnThrAsnAlaAlaLeuGlnAspValArgGluLeuLeuArg 20
QY 178 CAGCAGGTGAGGAGATGACGTTCTGGAAGAAACAGGAGATGAGTGGAGCGCGCGG 237
DB 21 GlnGlnValArgGluLeuThrPheLeuLysAsnThrValMetGluCysAspAlaCysGly 40
QY 238 CCGCAG-----CCGCAGCCGAAACCG 258
DB 41 MetGlnGlnSerValArgThrGlyLeuProSerValArgPro 54

RESULT 7
Q94885 PRELIMINARY; PRT; 1319 AA.
ID Q94885
AC Q94885;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA0790 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA MEDLINE=99087487; PubMed=9872452;
RX Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kocani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
```

```
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
DR EMBL; AB018333; BAA34510.1; -.
DR InterPro; IPR001600; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00536; SAM; 2.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00454; SAM; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 2.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 1319 AA; 143613 MW; FB97E509FEDBEP93 CRC64;

Alignment Scores:
Pred. No.: 7.33e-06 Length: 1319
Score: 153.00 Matches: 38
Percent Similarity: 50.47% Conservatve: 16
Best Local Similarity: 35.51% Mismatches: 27
Query Match: 22.77% Indels: 26
DB: Gaps: 5

US-09-696-872-24 (1-372) x Q94885 (1-1319)
QY 31 GCGTTGCTGCGCCCTTGGCGGACGCTGCGAGCTGCTGCGAGCTGCGAGCTGCGAGT 87
DB 23 GlyValAlaGlyAlaAlaGlyAlaCysSerValaGlyAlaGlyAlaGlyAlaGly 42
QY 88 GGAAGCTGTTTCTCA-----GACCTGGGCGCCGACGATGCTTGGGAACTG 132
DB 43 GlyAspProAlaSerGlyGlnAlaAlaArgGlyCysGlyAlaArgAlaProArgGlyLeu 62
QY 133 CAGAAACCAACGCGCGCGCTGAGAGAGCTGCGGACGCTGCGGACGAGTACAGGAG 192
DB 63 GlyArgThrAlaAlaGala----- 68
QY 193 ATCAGCTTCTGAAACACCGGTGATGAG--TGTGACGCGTCCGGGCGCAGCCGCGAG 249
DB 69 -----ArgAspThrAlaMetGlnAspAlaGlyAlaGlyAlaGlyProGlyProGlu 84
QY 250 CCGAAACCGCAGCGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAG 309
DB 85 ProGluProGluProGluProGluProGluProGluProGluProGluProGluProGlu 104
QY 310 ---GGTACCGGATCATCAGAA 327
DB 105 ProGlyAlaGlyThrSerGlu 111

RESULT 8
Q9QYS3 PRELIMINARY; PRT; 863 AA.
ID Q9QYS3
AC Q9QYS3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Thrombospondin 4 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Wietemowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thrombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF152393; AAD32714.1; -.
DR EMBL; AF152392; AAD32714.1; JOINED.
DR HSSP; P35444; IYDF.
MGD; MGI:1101779; Thbs4.
```

```

DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR001329; TSPN.
DR InterPro; IPR003367; TSP_3.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF02412; TSP_3; 9.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW EGF-like domain.
FT NON TER 1
SQ SEQUENCE 863 AA; 95339 MW; 68E3EE846728E4D CRC64;

Alignment Scores:
Pred. No.: 2,89e-05 Length: 863
Score: 146.50 Matches: 30
Percent Similarity: 51.85% Conservative: 12
Best Local Similarity: 37.04% Mismatches: 36
Query Match: 21.80% Indels: 3
DB: 11 Gaps: 1

US-09-696-872-24 (1-372) x Q9Y53 (1-863)

QY 67 AAAAAAGATCCAGCTGGGTGAGACTGTTTCAGACTGGGCCCGCAGATGCTTCGG 126
   : : : : :
DB 108 GlnGlnSerGluProLeuAlaAlaThrSerThrGlyAspPheAsnArgGlnPheLeuGly 127

QY 127 GAATGCGAGAAACCAACGCGCGCTGCGAGACGTGCGGAGCTGCGGAGCAGGTC 186
   : : : : :
DB 128 GlnMetThrGlnLeuAsnGlnLeuLeuGlyGluValIysAspLeuLeuArgGlnGlnVal 147

QY 187 AGGGAGATCAGCTTCCTGAAACACGCGTATGAGTGTGACGCGTGGCGGCGCAGCG 246
   : : : : :
DB 148 LysGluThrSerPheLeuArgAsnThrIleAlaGluCysGlnAlaCysGlyProLeuSer 167

QY 247 CAGCGGAAACCGCAGCCGCGCAG-----CCGACGCCGCGAGCGGAAACCGCAGCCGAAA 297
   : : : : :
DB 168 PheGlnSerProThrProAsnThrLeuValProIleAlaProProAlaProProThrArg 187

QY 298 CCG 300
   : : :
DB 188 Pro 188

RESULT 9
Q9Z1T2 PRELIMINARY; PRT; 963 AA.
AC Q9Z1T2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Thrombospondin-4.
GN THBS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Metemowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thrombospondin-4 gene."
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF102887; AAC73003.1; -
DR HSSP; P35444; 1VDF.
DR MGI; MGI:1101779; Thbs4.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF-like.

```

```

DR InterPro; IPR003129; TSPN.
DR InterPro; IPR003367; TSP_3.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF02412; TSP_3; 9.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW EGF-like domain.
SQ SEQUENCE 963 AA; 106366 MW; B8BA83B84F489FB1 CRC64;

Alignment Scores:
Pred. No.: 2.9e-05 Length: 963
Score: 146.50 Matches: 30
Percent Similarity: 51.85% Conservative: 12
Best Local Similarity: 37.04% Mismatches: 36
Query Match: 21.80% Indels: 3
DB: 11 Gaps: 1

US-09-696-872-24 (1-372) x Q9Z1T2 (1-963)

QY 67 AAAAAAGATCCAGCTGGGTGAGACTGTTTCAGACTGGGCCCGCAGATGCTTCGG 126
   : : : : :
DB 208 GlnGlnSerGluProLeuAlaAlaThrSerThrGlyAspPheAsnArgGlnPheLeuGly 227

QY 127 GAATGCGAGAAACCAACGCGCGCTGCGAGACGTGCGGAGCTGCGGAGCAGGTC 186
   : : : : :
DB 128 GlnMetThrGlnLeuAsnGlnLeuLeuGlyGluValIysAspLeuLeuArgGlnGlnVal 247

QY 187 AGGGAGATCAGCTTCCTGAAACACGCGTATGAGTGTGACGCGTGGCGGCGCAGCG 246
   : : : : :
DB 248 LysGluThrSerPheLeuArgAsnThrIleAlaGluCysGlnAlaCysGlyProLeuSer 267

QY 247 CAGCGGAAACCGCAGCCGCGCAG-----CCGACGCCGCGAGCGGAAACCGCAGCCGAAA 297
   : : : : :
DB 268 PheGlnSerProThrProAsnThrLeuValProIleAlaProProAlaProProThrArg 287

QY 298 CCG 300
   : : :
DB 288 Pro 288

RESULT 10
Q25648 PRELIMINARY; PRT; 272 AA.
AC Q25648;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Circumsporozoite (CS) protein (Fragment).
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5821;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87218962; PubMed=3556207;
RA Weber J.L., Egan J.B., Lyon J.A., Wirtz R.A., Charoenvit Y.,
RA Maloy W.L., Hockmeyer W.T.;
RT "Plasmodium berghei: Cloning of the circumsporozoite protein gene."
RL Exp. Parasitol. 63:295-300(1987).
DR EMBL; M25445; AA829531.1; -
DR InterPro; IPR003067; Circmepizoite.
DR InterPro; IPR002965; P_fich_extensn.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; TSP_1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00092; TSP1; 1.
FT NON TER 1
SQ SEQUENCE 272 AA; 29408 MW; 4FF07FA62B32A051 CRC64;

```


This Page Blank (uspio)